



Structures et Fonctions des séquences subtélomériques productrices de piRNA

Amna Asif-Laidin

► To cite this version:

Amna Asif-Laidin. Structures et Fonctions des séquences subtélomériques productrices de piRNA. Génétique. Université Pierre et Marie Curie - Paris VI, 2016. Français. NNT : 2016PA066014 . tel-01376943

HAL Id: tel-01376943

<https://theses.hal.science/tel-01376943>

Submitted on 6 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

Ecole doctorale Complexité du Vivant, ED 515

IBPS - Laboratoire de Biologie du Développement UMR7622-CNRS

Université Pierre et Marie Curie

Equipe : Répression épigénétique et ADN mobile

Structures et Fonctions des séquences subtélomériques productrices de piRNA

Par Amna Asif-Laidin

Thèse de doctorat de Génétique

Dirigée par Laure Teyssset

Présentée et soutenue publiquement le 04 avril 2016

Devant un jury composé de :

Dr. Sandra Duharcourt	DR2, CNRS	Rapporteur
Pr. Cristina Vieira	Professeur, Université Claude Bernard Lyon 1	Rapporteur
Pr. Frédéric Devaux	Professeur, UPMC	Examineur
Dr. Jean René Huynh	DR2, CNRS	Examineur
Dr. Wolfgang Miller	Assistant, Professor University of Vienna	Examineur
Dr. Laure Teyssset	Maître de Conférences, UPMC	Directeur de thèse

Introduction générale

« The fly room » est le nom qui a été donné au laboratoire de Thomas Morgan Hunt voilà maintenant plus d'un siècle. Il s'agit de la pièce où les drosophiles étaient étudiées et où il vu pour la première fois une mouche mutante aux yeux blancs parmi une population de mouches sauvages aux yeux rouges. Ses travaux sur l'hérédité chromosomique lui ont permis de recevoir le prix Nobel de médecine en 1933. De nos jours, une « pièce à mouches » est présente dans chaque laboratoire étudiant la drosophile. Cet organisme modèle est un puissant outil pour l'étude de nombreux processus génétiques et développementaux. En effet, la drosophile a une très petite taille (3mm), son cycle de vie est relativement court (environ 12 jours à 25°C) et on peut obtenir un grand nombre descendants. Elle possède 4 paires de chromosomes et son génome a été entièrement séquencé (Adams *et al.* 2000). Une base de données (flybase.org) est aussi disponible regroupant un très grand nombre d'informations sur les gènes, les allèles mutants, les phénotypes etc. Au laboratoire, nous utilisons la drosophile afin de comprendre les mécanismes épigénétiques mis en jeu dans la protection de la transmission de l'information génétique au cours des générations. Nous étudions cela à travers la répression des éléments transposables (ET) dans la lignée germinale car c'est dans ces tissus que les ET sont actifs pouvant causer des mutations sur l'ADN ainsi que des réarrangements chromosomiques. Je vous introduirai donc dans un premier temps les éléments transposables puis je vous détaillerai plus particulièrement la région télomérique de *D. melanogaster* impliquée dans la répression d'un ET que nous étudions au laboratoire, l'élément *P*. Je finirai par la description des mécanismes responsables de cette répression connus à l'heure actuelle.

Introduction générale	3
I-Synthèse bibliographique	8
Chapitre 1 : Les éléments transposables	9
A. Description de la lignée germinale	9
B. Classification des éléments transposables	11
B.1 Les rétrotransposons (classe I)	12
B.2 Les transposons (classe II)	13
B.3 La localisation des ET dans les génomes	15
B.4 L'impact des ET sur les génomes	16
C. Le cas de l'élément <i>P</i> chez la drosophile	17
C.1 Présentation de l'élément <i>P</i>	17
C.2 La dysgénésie des hybrides et la régulation de l'élément <i>P</i>	18
D. Utilisation de transgènes <i>P-lacZ</i> pour l'étude de la régulation de l'élément <i>P</i> : Le <i>Trans-Silencing Effect</i>	20
D.1 La répression par les <i>P-lacZ</i> : les transgènes silenciers et cibles	20
D.2 Les propriétés du TSE	21
D.4 Les acteurs moléculaires du TSE	24
Chapitre 2 : Organisation des régions télomériques chez <i>D. melanogaster</i>	26
A. Les TAS	26
A.1 Structure des TAS	27
A.1.1 La famille des TAS-R	27
A.1.2 La famille des TAS-L	29
A.1.3 Le cas du chromosome 4 et du chromosome Y	29
A.2 Les protéines associées aux TAS	30
A.3 Le rôle des TAS	33
B. Les rétrotransposons télomériques : <i>HeT-A</i> , <i>TAHRE</i> et <i>TART</i>	34
B1. <i>HeT-A</i> , <i>TART</i> et <i>TAHRE</i>	35
B2. Mécanisme de transposition et élongation du télomère	35
C. La cap	36
Chapitre 3 : Les petits ARNs impliqués dans la régulation des ET en lignée germinale	39
A. Les PIWI-interacting RNA (piRNA)	39
A.1 Les clusters producteurs de piRNA	39
A.2 La biogenèse des piRNA germinaux	40
A.2.1 Les piRNA germinaux primaires	43
A.2.2 Les piRNA germinaux secondaires	44
A.2.3 Les piRNA germinaux tertiaires	46
A.3 La biogenèse des piRNA folliculaires somatiques	46
A.4 La répression des éléments transposables par les piRNA	49
A.4.1 répression transcriptionnelle : TGS	49
A.4.2 répression post-transcriptionnelle : PTGS	50
A.5 L'activation d'un cluster pour la production de piRNA <i>de novo</i>	51
B. Régulation de gènes par les piRNA	51
B.1 Les 3'UTR, une source de piRNA dans les cellules somatiques	52
B.2 La régulation du gène <i>nanos</i> par les piRNA	53
B3. Régulation de gènes par les piRNA chez d'autres espèces	Erreur ! Signet non défini. 55
B.1.1 La souris	55
B.1.2 Le ver à soie	56

B.1.3 L'aplysie	57
C. Objectifs de l'étude	58
II-Résultats.....	60
1. Article I.....	61
Présentation l'article I	61
Article I. <i>Telomeric Associated Sequences in Drosophila: structure, dynamic and properties of a piRNA cluster</i>	62
Conclusions	109
2. Analyse de mutants pour la production de piRNA du locus TAS	111
A1. Les TAS-R.....	111
A2. Les TAS-L.....	112
3. Article II.....	115
Présentation de l'article II	115
Article II. <i>Paramutation in Drosophila Requires Both Nuclear and Cytoplasmic Actors of the piRNA Pathway and Induces Cis-spreading of piRNA Production</i>	116
Conclusions	148
4. Article III	150
Présentation de l'article III.....	150
Article III. <i>Reprogramming of a sequence fate after its insertion into a piRNA cluster</i>	151
Conclusion.....	196
5. Régulation de l'expression du génome par les piRNA du locus TAS	198
A. Hypothèse de travail	198
B. Recherche de gènes candidats potentiellement ciblés par les TAS-R	199
B1. Analyse <i>in silico</i>	200
B2. Les séquences TAS intra-géniques.....	201
B3. Les séquences TAS flanquantes aux gènes	202
C. Analyse par qRT-PCR	203
D. Construction de lignées sans TAS.....	205
E. Les piRNA du domaine F peuvent-ils éteindre une cible ?	206
E1. Construction des transgènes et descriptions des lignées utilisées.....	206
E2. Validation <i>in vivo</i> des lignées <i>pUASp-RFP-DomaineF</i>	207
E2.1 Principe de l'expérience.....	207
E2.2. Expression du transgène <i>pUASp-RFP-K10</i> sans le domaine F.....	209
E2.2.1 Absence de piRNA du domaine F	209
E2.2.2 Avec piRNA du domaine F.....	209
E2.3. Expression du transgène <i>pUASp-DomF-RFP-K10</i>	211
E2.3.1 Absence de piRNA du domaine F	211
E2.3.2 Avec piRNA du domaine F.....	214
E2.4 Le TSE par une cible P-lacZ.....	215
F3. Rôle des piRNA du <i>TAS-X</i> dans l'expression du génome.	217
F.3.1 Analyse du transcriptome des lignées isogénisées.....	218
F.3.1.1 Présentation des lignées	218
F.3.1.2 Analyse préliminaire	219
III- Discussion et perspectives.....	224
IV-Matériels et Méthodes	233

A-Hybridation <i>in situ</i> sur chromosomes polytènes	234
A1. Préparation des squashes et de la sonde ADN.	234
A2. Hybridation <i>in situ</i> avec sonde ADN biotinylée.....	236
Pré-hybridation.....	236
Dénaturation	236
Préparation de la solution d'hybridation	236
Hybridation.....	237
Détection et révélation.....	237
A3. Hybridation <i>in situ</i> avec sonde ADN fluorescente	238
Préparation de la sonde.....	239
Hybridation.....	239
Détection et révélation.....	239
B. Analyse des petits ARN.....	252
C. Détection de la fluorescence endogène en ovaires de drosophile.....	253
V-Références bibliographiques	255
VI-Annexes.....	256

I-Synthèse bibliographique

Chapitre 1 : Les éléments transposables

A. Description de la lignée germinale

L'ovogenèse chez la drosophile est un processus dynamique et constant. La femelle possède la capacité remarquable de générer un large nombre d'œufs durant toute sa vie grâce à l'existence permanente de cellules souches.

La femelle possède une paire d'ovaires, qui sont eux mêmes composés d'environ une quinzaine d'ovarioles (Figure 1). Chaque ovariole est une succession de chambres à œuf (ou follicule ovarien) à différents stades de développement organisée linéairement selon un axe antéro-postérieur, la chambre à œuf la plus mature étant située du côté postérieur.

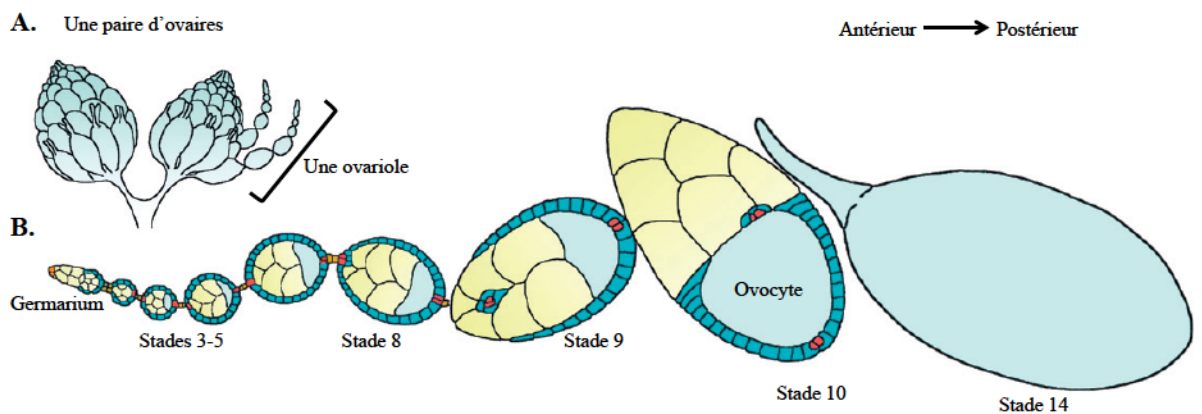


Figure 1. A. Représentation schématique d'une paire d'ovaire Une paire d'ovaires est constituée de plusieurs ovarioles. **B. Représentation schématique d'une ovariole.** Dans une ovariole, on retrouve des chambres à œufs qui possèdent un degré de maturité croissante selon un axe antéro-postérieur (indiqué par une flèche). En bleu foncé, on retrouve les cellules folliculaires somatiques et en jaune, les cellules germinales (He *et al.* 2011)

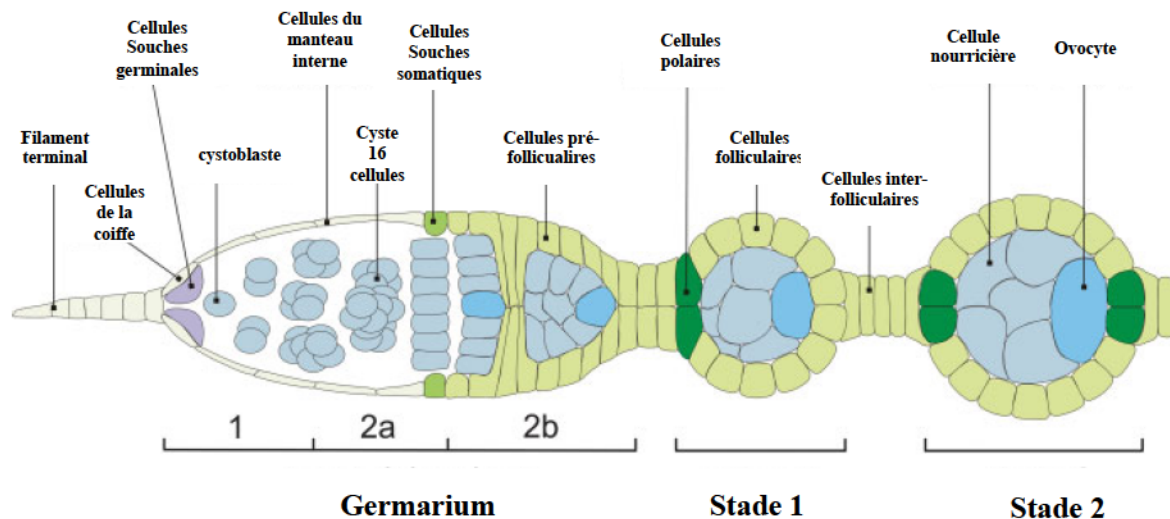


Figure 2. Représentation schématique du germaireum. (D'après Horne-Badovinac *et al.* 2005). Le germaireum contient de deux grandes populations de cellules souches, les cellules souches germinales (ou GSC « germline stem cell ») et les cellules souches somatiques (ou SSC « somatic stem cell »).

A la partie, la plus antérieure on trouve la structure appelée germaireum qui se sous divise en 4 régions : 1, 2a 2b et 3 (la région 3 correspondant au stade 1) (Figure 2). Il contient deux types de cellules souches : les cellules souches germinales (GSCs « germline stem cells ») et les cellules souches somatiques (SSCs « somatic stem cells»). Dans la région 1, les GSCs se divisent asymétriquement pour engendrer une cellule souche permettant une régénération des GSCs et une cellule fille. La cellule fille subit alors quatre divisions mitotiques mais avec une cytokinèse incomplète formant ainsi, dans la région 2a, un cyste de 16 cellules interconnectées entre elles par un pont cytoplasmique. Parmi ces 16 cellules, l'une d'entre elle se différenciera en ovocyte tandis que les 15 cellules restantes deviendront des cellules nourricières polyploïdes. A la frontière de la région 2a et 2b, on retrouve les SSCs qui produisent les cellules pré-folliculaires. Ces cellules vont alors encapsuler le cyste lui conférant l'identité du premier follicule ovarien dans la région 3 (stade 1 figure 2) (He *et al.* 2011) (Horne-Badovinac *et al.* 2005) .

On peut distinguer les différentes phases du développement du follicule ovarien selon 14 stades prédéfinis par Robert C. King. (Figure 1). Le stade 1 étant le follicule ovarien qui bourgeonne du germaireum alors que le stade 14 correspond au follicule le plus mature prêt à être pondu. On distingue les stades pré-vitellogénique (1 à 7) au cours desquels toutes les cellules germinales (cellules nourricières et ovocyte) ont à peu près la même taille et les

stades vitellogéniques à partir du stade 8, au cours desquels la taille de l'ovocyte va augmenter considérablement par rapport aux cellules nourricières qui produisent les ARN et les protéines maternels nécessaires au bon développement de l'ovocyte.

B. Classification des éléments transposables

« Jumping genes » autrement dit « gènes sauteurs » est le nom qui a été attribué aux éléments transposables suite à leur découverte chez le maïs par Barbara McClintock il y a plus de 60 ans. Cette avancée lui a permis d'obtenir le prix Nobel de médecine en 1983. Un élément transposable (ET) ou transposon est donc une séquence d'ADN mobile qui est capable de se déplacer de manière autonome dans le génome par un mécanisme appelé « transposition ». On retrouve des ET dans toutes les grandes branches du monde vivant même si leur proportion peut-être variable selon les génomes (Figure 3). Chez l'abeille, *Apis mellifera*, les ET représente 1% du génome, chez la mouche du vinaigre, *Drosophila melanogaster* 20%, 45% chez l'homme et ce chiffre peut atteindre presque les 70% chez certaines espèces comme le maïs, *Zae mays* (Biemont 2010). Selon leur structure et leur molécule intermédiaire de transposition (ARN ou ADN), deux grandes familles d'ET ont pu être définies par Finnegan en 1989. On distingue donc les rétrotransposons et les transposons.

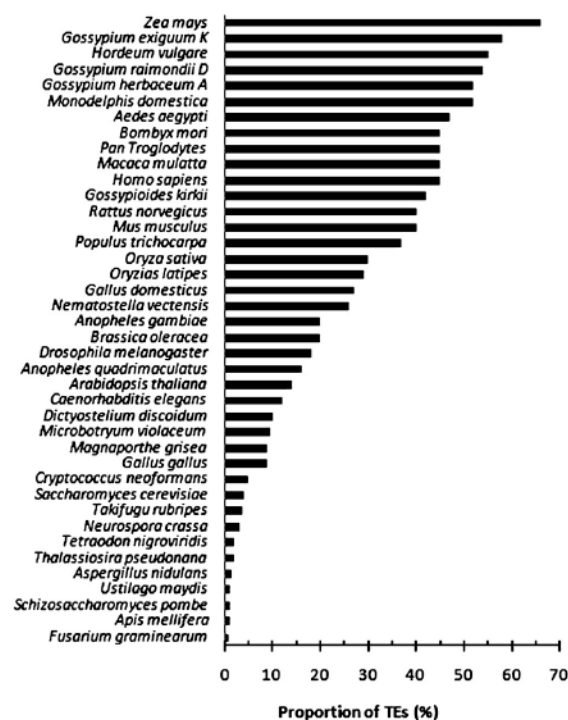


Figure 3. Proportion d'ET chez différentes espèces eucaryotes (Biemont 2010)

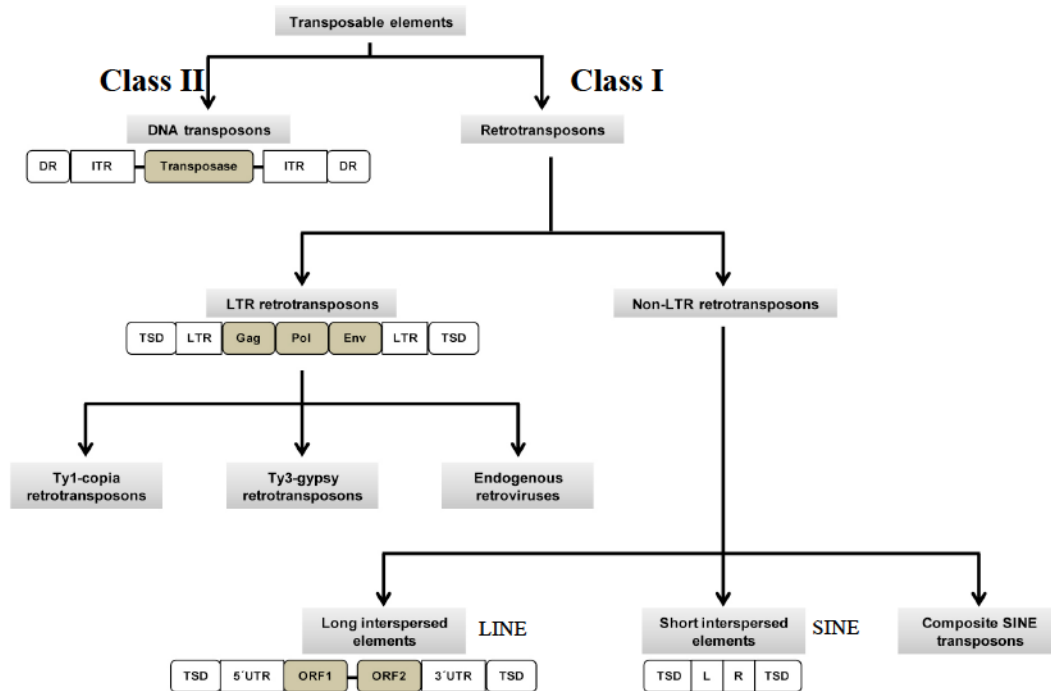


Figure 4. Classification et structure des ET DR : direct repeat ; ITR : inverted terminal repeat ; TDS : target site duplication; LTR : long terminal repeat ; UTR : untranslated region ; ORF : open reading frame ; L : Left ; R : right ; SINE : short interspersed nuclear element (Ayarpadikannan and Kim 2014).

B.1 Les rétrotransposons (classe I)

Les éléments de la classe 1 ou rétrotransposons ont la particularité de transposer *via* un intermédiaire ARN grâce à l'activité d'une transcriptase inverse. Ces éléments fonctionnent selon un mécanisme de « copier-coller », ils sont d'abord rétro-transcrits afin de pouvoir se réinsérer dans l'ADN génomique (Figure 5). Ceci permet l'augmentation de leur nombre de copies. Une marque caractéristique de l'insertion des ET dans le génome est le TDS (Target Site Duplication). C'est une séquence de 4 à 8pb que l'on retrouve de part et d'autre du site d'insertion de l'ET (Figure 4). On distingue deux grandes catégories parmi les rétrotransposons selon la présence ou non d'une structure appelée LTR (Long Terminal Repeat) qui sont de longues répétitions terminales (Figure 4).

Les rétrotransposons à LTR possèdent deux LTR qui sont entourés par des cadres ouverts de lecture (ou ORF pour « Open Reading Frame ») codant pour *gag*, *pol* et plus exceptionnellement pour *env* rappelant ainsi la structure des rétrovirus (Kim *et al.* 1994, Teyssset *et al.* 1998). Cette classe se subdivise à nouveau en plusieurs super-familles notamment les *Ty1/copia* et *Ty3/gypsy* qui diffèrent par la position de l'intégrase au sein de *pol*. Les *LINE* (Long Interspersed Element) et les *SINE* (Short Interspersed Nuclear Element)

sont deux super-familles faisant partie des rétrotransposons sans LTR. Les *LINE* sont des éléments possédant deux ORF dont l'une code au moins une reverse transcriptase pour leur transposition. Ils sont également caractérisés par la présence d'une queue poly-A ou bien d'une région riche en A à leur extrémité 3'. Les *SINE* sont des éléments non autonomes car ils utilisent la reverse transcriptase des *LINE* afin de pouvoir transposer. Ces éléments possèdent un promoteur interne de l'ARN pol III permettant leur expression. Les régions internes des *SINE* (L et R, voir figure 4) peuvent être très variables selon leurs sous-familles. Un *SINE* très connu est l'élément *Alu* présent chez l'homme et représentant 10,6% de son génome (Cordaux and Batzer 2009).

B.2 Les transposons (classe II)

Les éléments de classe II transposent par un intermédiaire ADN par un mécanisme dit de « couper-coller » (Figure 5). Ils sont caractérisés par la présence d'une ORF codant pour la transposase qui est entourée de répétitions terminales inversées appelées *ITR* pour « *Inverted Terminal Repeat* » dont la taille peut-être variable selon les éléments (Wicker *et al.* 2007). Ces ITR permettent à la transposase de pouvoir se fixer et aider à la transposition de l'élément. Ils sont donc directement excisés puis réintégrés ailleurs dans le génome. Parmi ces transposons on retrouve l'élément *P* présent chez *Drosophila melanogaster*.

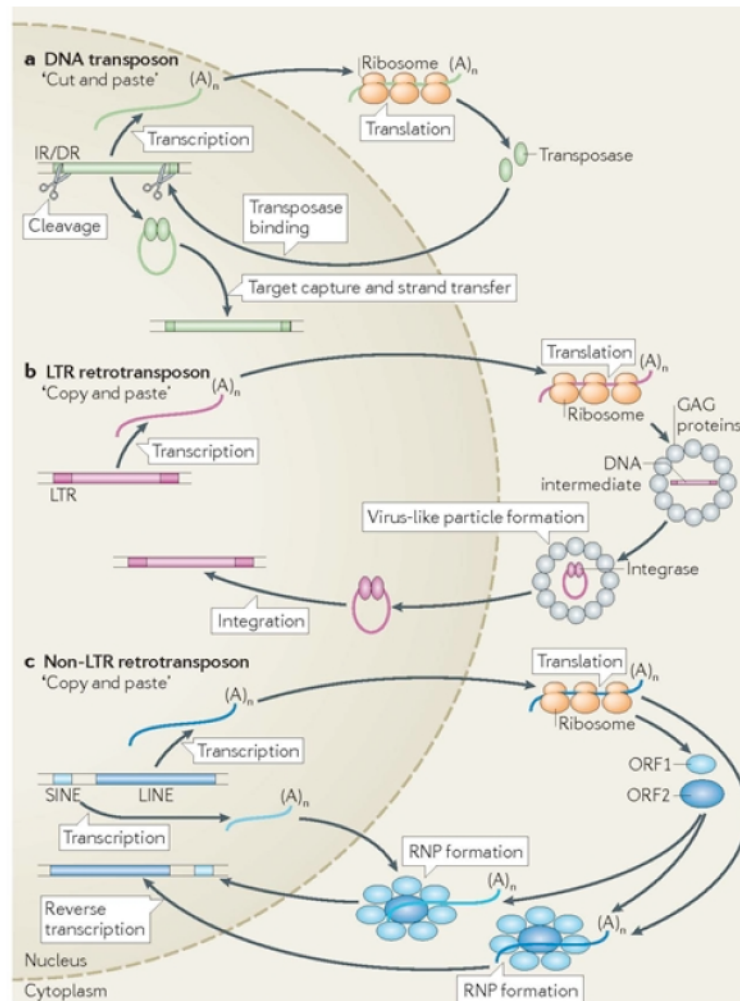


Figure 5. Cycle de transposition des différentes classes d'éléments transposables. A. En vert, les transposons transposent par un mécanisme dit de « couper-coller ». Ces transposons codent pour une transposase qui va être capable de les exciser et les réintégrer ailleurs dans le génome. **B.** En rose, les rétrotransposons transposent via un mécanisme de « copier-coller ». Ils codent pour la GAG et l'intégrase. Les protéines GAG peuvent former une particule virale où à lieu la reverse transcription de l'ARNm. L'intermédiaire ADN est transporté jusqu'au noyau où il est réintégré dans le génome **C.** En bleu, les rétrotransposons sans LTR transposent aussi via un mécanisme de « copier-coller ». Les LINE codent pour l'ORF 1 et l'ORF 2. Ces protéines forment un complexe RNP avec leur ARNm qui sont transportés jusqu'au noyau où a lieu une reverse transcription puis l'intégration dans le génome. Les SINE utilisent la machinerie des LINE pour leur reverse transcription. ORF : Open Reading Frame, IR : Inverted Repeats, DR : Direct repeats. RNP : « Ribonucleoprotein » (Siomi *et al.* 2011).

B.3 La localisation des ET dans les génomes

Par définition, les ET étant des éléments mobiles, on les retrouve dispersés un peu partout au sein des génomes. Cependant, il semble exister des régions chromosomiques où l'on peut observer une densité plus importante de ces éléments. En effet, les régions péri-centromériques et les télomériques sont plus riches en ET que le reste des bras chromosomiques par exemple chez *Drosophila melanogaster* (Bergman *et al.* 2006). Ces régions sont pour la majorité des régions hétérochromatiques pauvres en gènes (Figure 6). En effet, 66% de ces régions sont composées de rétrotransposons (Smith *et al.* 2007)

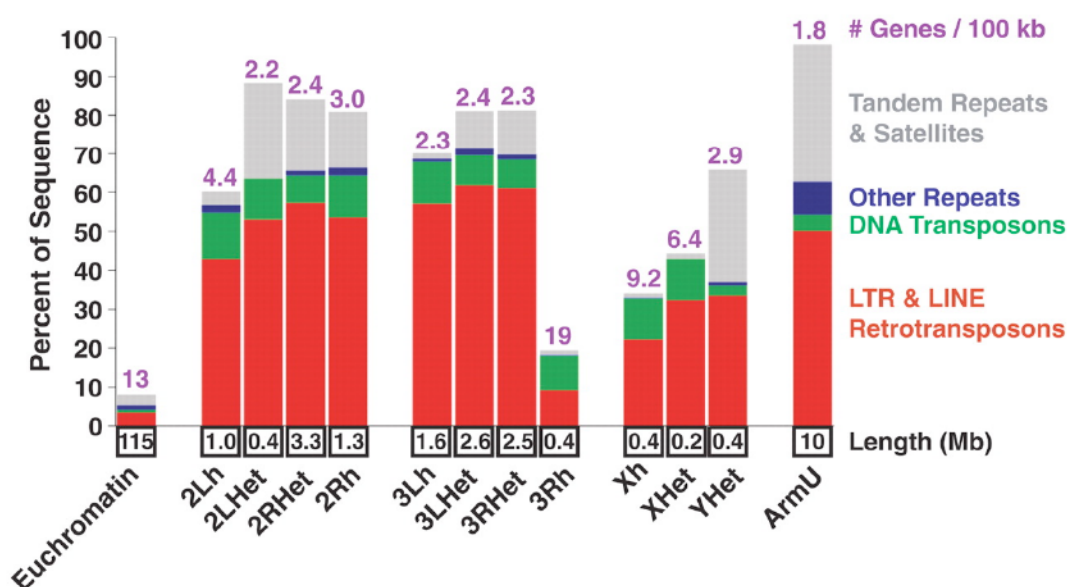


Figure 6. Densité de l'ADN répété et des gènes au sein de l'hétérochromatine. Le graphique représente le pourcentage de chaque type d'annotation par chromosome. L'euchromatine regroupe les chromosomes 2, 3 et X. Les chromosomes 2Lh, 2Rh, 3Lh, 3Rh, et Xh sont des contigs de régions hétérochromatiques non annotés alors que les chromosomes appelés Het ont pu être annotés. On retrouve, en rouge le pourcentage de rétrotransposons à LTR ainsi que de LINE; En vert, les transposons à ADN; en bleu les répétitions d'origines inconnues et en gris les répétitions en tandem ainsi que les séquences satellites. Le nombre moyen de gènes par 100kb est représenté pour chaque histogramme (Smith *et al.* 2007).

B.4 L'impact des ET sur les génomes

En 1972, Susumu Ohno publie un article intitulé « so much Junk DNA in our genome », regroupant sous ce terme un grand nombre de séquences présentes sans réelle fonction apparente dans les génomes dont les ET. Cet ADN a ensuite été catégorisé comme étant de « l'ADN égoïste » ayant pour simple but d'envahir le génome en augmentant son nombre de copies et ne contribuant pas au phénotype. Cette théorie a été émise lors de deux publications par Orgel et Crick ainsi que Doolittle et Sapienza dans la même édition de la revue *Nature* en 1980 (Doolittle and Sapienza 1980), (Orgel and Crick 1980). Cependant, les auteurs laissent tout de même la porte ouverte quant à l'éventuel rôle de cet ADN égoïste « Nous ne nions pas que les ET procaryotes ou les séquences répétées ainsi que les séquences uniques d'ADN ne codant pas pour les protéines chez les eucaryotes pourraient avoir un rôle bénéfique immédiat sur le phénotype de l'organisme. Cependant nous ne nions pas le rôle de ces éléments dans le processus de l'évolution » (Doolittle and Sapienza 1980)

Les éléments transposables sont-ils alors un bénéfice ou une contrainte pour le génome ? La réponse à cette question est très complexe car les ET peuvent à la fois avoir des effets négatifs ou positifs dans la cellule (pour revue, (Feschotte and Pritham 2007);(Biemont and Vieira 2006);(Rebollo *et al.* 2012)). Par exemple, certaines maladies sont associées aux ET. En effet, une insertion d'un LINE dans le gène codant pour le facteur VIII est responsable de l'hémophilie. Ainsi plus de 60 insertions *de novo* d'ET ont été répertoriées pour leur implication dans des maladies génétiques chez l'homme (Cordaux and Batzer 2009). Par ailleurs, il existe un exemple très connu de domestication moléculaire dans la recombinaison V(D)J jouant un rôle dans le système immunitaire des vertébrés. Ce processus permet de générer une grande diversité d'anticorps dans les lymphocytes B et T. Les deux protéines RAG1 et RAG2 sont les deux composants permettant cette recombinaison qui dériveraient de la transposase d'un transposon à ADN de la famille *Transib*. Un autre cas de domestication est celui des éléments *HeT-A*, *TART* et *TAHRE* qui transposent spécifiquement aux télomères et remplissent le rôle de la télomérase chez la drosophile (voir chapitre 2). Enfin, les ET peuvent aussi être des outils intéressants pour la manipulation des génomes. Par exemple l'utilisation de l'élément *P* a permis de faire de la transgénèse chez la drosophile car celui-ci peut s'insérer plus ou moins au hasard dans le génome.

C. Le cas de l'élément *P* chez la drosophile

C.1 Présentation de l'élément *P*

L'élément *P* est un transposon à ADN qui a été très étudié dans les laboratoires car il a envahi le génome des populations naturelles de *Drosophila melanogaster* au cours du siècle dernier. Il est impliqué dans le phénomène de dysgénésie des hybrides (voir partie suivante). Ce transposon de 2,9kb possède quatre exons codants pour la transposase, flanqués d'ITR de 31pb (O'Hare and Rubin 1983). Cet élément code pour deux types de protéines : la transposase de 87kDa présente en lignée germinale ainsi qu'une version tronquée de celle-ci, un répresseur de la transposition de 66kDa pouvant être présent en lignée germinale et somatique (Misra and Rio 1990) (Figure 7). Ceci est possible grâce à un épissage alternatif qui peut avoir lieu entre les exons 2 et 3 dans les tissus somatiques. On peut dénombrer plus de 50 copies d'élément *P* dans certaines souches de drosophiles comme par exemple dans la souche « Harwich » (Ronsseray *et al.* 1997).

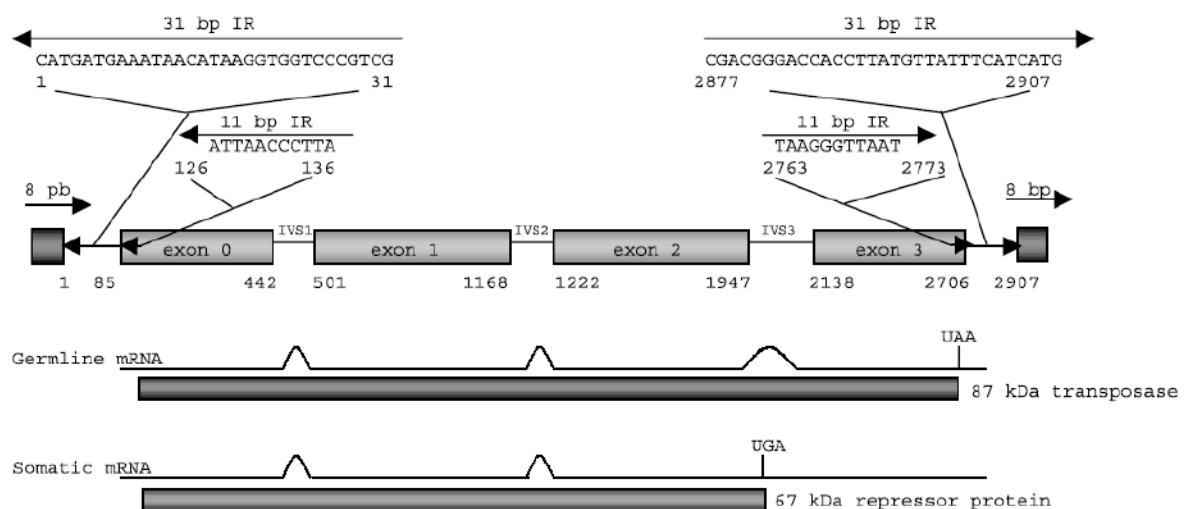


Figure 7. Structure de l'élément *P* On retrouve les ITR de 31pb appelés ici IR entourant les 4 exons codants la transposase. Deux transcrits ont été décrits à partir de cet élément, un ARNm spécifique de la lignée germinale codant pour la transposase de 87kDa et un ARNm spécifique du soma codant pour le répresseur de 66kDa (de Castro and Carareto 2004).

C.2 La dysgénésie des hybrides et la régulation de l'élément *P*

Au début des années 60, les généticiens des populations ont observés des effets délétères dans la descendance de croisements entre des drosophiles récemment récoltées dans la nature et des drosophiles conservées dans les laboratoires (Hiraizumi and Crow 1960). Plus tard, ces observations ont été regroupées sous le terme « dysgénésie des hybrides » (Kidwell *et al.* 1977) qui sont un ensemble d'anomalies génétiques résultant de l'activité de transposition des éléments *P*. D'autres ET peuvent aussi induire de la dysgénésie des hybrides tels que le facteur *I* et *hobo*. Ce phénomène a lieu dans la lignée germinale et se traduit par un ensemble de symptômes tels que : des cassures de l'ADN, des réarrangements chromosomiques, un fort taux de mutation ainsi qu'une stérilité thermo-dépendante de la descendance due à une atrophie des gonades à 29°C (Kidwell *et al.* 1977) (Engels 1979). Ce phénomène peut-être quantifié et permet de mesurer le taux de l'activité *P* dans une souche donnée en comptant le nombre de paire d'ovaires atrophiés. La dysgénésie des hybrides a lieu quand on croise une femelle ne portant pas d'élément *P* (appelée souche M) par un mâle possédant des éléments *P* (appelée souche P) dans son génome. On observe alors une forte activité *P*. Par exemple quand on croise une femelle *w*¹¹¹⁸ (souche M) par un male Harwich (souche P), on peut avoir 100% d'activité *P*. A l'inverse lorsqu'on effectue le croisement réciproque, c'est à dire lorsqu'on croise une femelle P par un mâle M, cette dysgénésie n'a pas lieu (Figure 8). Ainsi la femelle P porte une information qu'elle transmet à sa descendance qui la protège contre la transposition des éléments *P*. Cet état de répression a été appelé le cytotype P (Engels 1979). La régulation de la transposition de l'élément *P* est donc soumise à un effet maternel.

On peut trouver des éléments *P* dispersés partout dans le génome cependant il existe un point chaud d'insertion au site cytologique 1A du chromosome *X* qui est une région télomérique (Ronsseray *et al.* 1989). Il a été montré qu'une ou deux copies de *P* insérées au site 1A sont capable de réprimer la transposition de plus d'une dizaine de copies présentes dans le génome (Ronsseray *et al.* 1991) (Marin *et al.* 2000). Quels sont les mécanismes permettant d'établir cette répression ? L'implication de l'hétérochromatine ainsi que du système d'ARN interférence (ARNi) dans la répression de l'élément *P* a été suggérée dans les années 90. En effet, il a été montré que la répression des *P* télomériques est sensible à la dose de HP1 (Heterochromatin Protein 1) (Ronsseray *et al.* 1996), (Marin *et al.* 2000). D'autre part, cette répression des éléments *P* télomériques est aussi affectée dans un contexte mutant pour le gène *Aub* (*Aubergine*) qui fait partie de la machinerie d'ARNi (Reiss *et al.* 2004). Ces données convergent vers un mécanisme faisant intervenir à la fois des acteurs nécessaires à la

formation de l'hétérochromatine et un système d'ARNi, tout en soulignant l'importance de la présence de *P* au site télomérique 1A. Je vous décrirai plus tard en détails comment se compose le site 1A ainsi que son rôle fonctionnel.

L'arrivée des techniques de séquençage à haut débit a permis de montrer que cette répression est effectuée par une classe spécifique de petits ARN non codants, les piRNA pour PIWI-interacting RNA (voir chapitre 3). Ces piRNA sont la composante cytoplasmique principale associée à l'effet maternel observé lors de la dysgénésie des hybrides (Figure 8) (Brennecke *et al.* 2008).

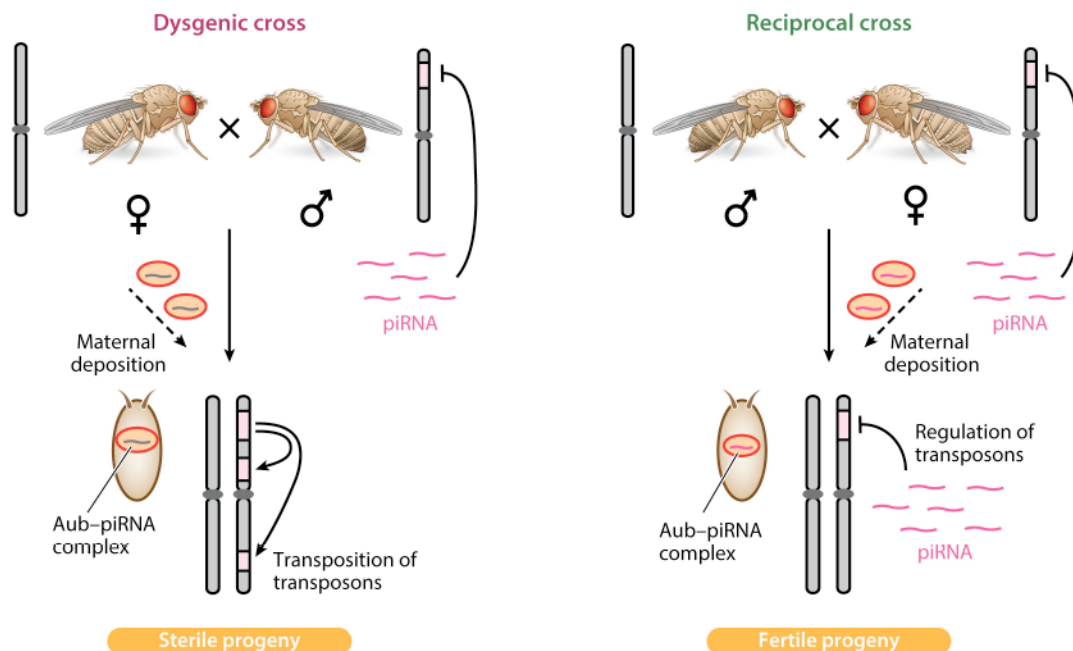


Figure 8. La dysgénésie des hybrides. A gauche est représenté un croisement dysgénésique, une femelle dépourvue d'éléments *P* est croisée par un mâle possédant cet ET. La descendance est stérile à cause de la transposition de cet ET dans le génome. A l'inverse, à droite, est représenté un croisement non dysgénésique. La femelle transmet à sa descendance des piRNA homologues à l'ET permettant de réprimer sa transposition. Dans l'ovale en rouge, on retrouve des complexes piRNA chargés par leurs protéines ici Aub pour Aubergine (voir chapitre 3) (Iwasaki *et al.* 2015).

Lors de l'étude de la régulation de la transposition de l'élément *P*, un mécanisme de répression homologie-dépendant a été découvert et appelé TSE pour *Trans-Silencing Effect*. Ceci a fait l'objet de nombreuses études au laboratoire.

D. Utilisation de transgènes *P-lacZ* pour l'étude de la régulation de l'élément *P* : Le *Trans-Silencing Effect*

Parmi les mécanismes connus de répression de l'expression du génome, il a été montré un mécanisme de répression spécifique de la lignée germinale chez *Drosophila melanogaster* appelé *Trans-Silencing Effect* (TSE) (Roche and Rio 1998). Ce processus met en jeu des transgènes *P-lacZ* télomériques et euchromatiques afin de mimer la répression de l'élément *P*. En effet, lorsqu'un transgène est inséré au site cytotologique 1A celui-ci est capable de réprimer en *trans* l'expression germinale d'un transgène homologue inséré dans les régions euchromatiques comme dans le cas de la répression de l'élément *P*.

D.1 La répression par les *P-lacZ* : les transgènes silencer et cible

Les transgènes *P-lacZ* sont des enhancer-trap. On trouve dans ces transgènes, le gène *lacZ* d'*E. coli* en fusion transcriptionnelle avec les deux premiers exons de l'élément *P*. Ces transgènes sont d'excellents outils pour l'étude de l'expression des ET en lignée germinale chez la drosophile. Après une coloration des ovaires au X-gal, si le gène *lacZ* est exprimé alors les ovaires sont bleus. A l'inverse si le gène est réprimé, les ovaires sont blancs.

Le TSE met en jeu deux types de transgènes, un transgène *P-lacZ* silencer et un transgène cible *P-lacZ* possédants des caractéristiques différentes.

-Les transgènes appelés « silencer » sont insérés dans les régions sub-télomériques hétérochromatiques appelées Telomere Associated Sequence (TAS) et sont la source de la composante maternelle permettant d'observer une répression. Après une coloration au X-gal d'une femelle portant un transgène silencer, on n'observe aucune coloration des ovaires (Figure 9A). Au laboratoire, nous travaillons sur différents transgènes silenciers parmi lesquels on retrouve *P-1152* possédant deux transgènes *P-lacZ* insérés au site 1A du chromosome X.

-Les transgènes appelés « cible » (*BC69*, *BQ16*, *PCo1*, *PboL*...) sont quant à eux localisés dans les régions euchromatiques. Après une coloration au X-gal d'une femelle portant un transgène cible, on observe une coloration bleue des ovaires (Figure 9B).

Le TSE peut être estimé en comptant le nombre de chambres blanches et le nombre de chambres bleues au sein d'un ovaire (Figure 10) ceci permet d'établir un pourcentage de répression.

A. *P-lacZ* cible



B. *P-lacZ* silencer



Figure 9. Expression des transgènes *P-lacZ* dans les ovaires A. L'expression du transgène euchromatique *P-lacZ* cible. On peut distinguer les noyaux des cellules nourricières des chambres à œufs ainsi que de l'ovocyte au stade 10 colorés en bleu après révélation au X-gal. B. On n'observe aucune expression du transgène silencer inséré dans les TAS (Josse *et al.* 2007).

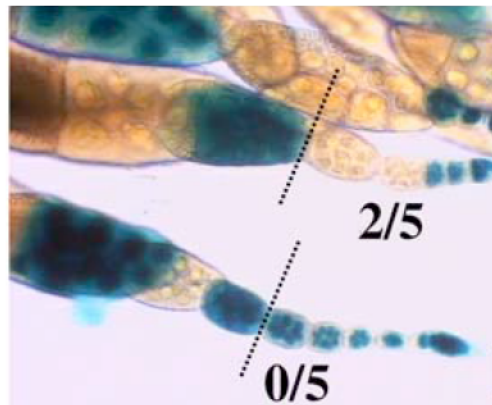


Figure 10. Quantification du TSE. Pour la quantification du TSE (pourcentage de répression), on compte le nombre de chambres réprimées parmi les cinq premiers stades. Dans cet exemple, il y a respectivement 2 et 0 chambres blanches sur 5 pour les deux ovarioles. Le pourcentage de TSE sera alors de $(2+0)/10 \times 100 = 20\%$. Afin d'avoir une estimation la plus proche de la réalité, il est important d'avoir un effectif élevé. Ainsi pour chaque expérience dans la mesure du possible, deux colorations lacZ indépendantes sont réalisées, nous disséquons 8 à 10 paires d'ovaires et 10 ovarioles ($n \geq 1000$) sont comptées pour chaque ovaires afin d'établir le pourcentage de TSE (Josse *et al.* 2007).

D.2 Les propriétés du TSE

Restreint à la lignée germinale femelle

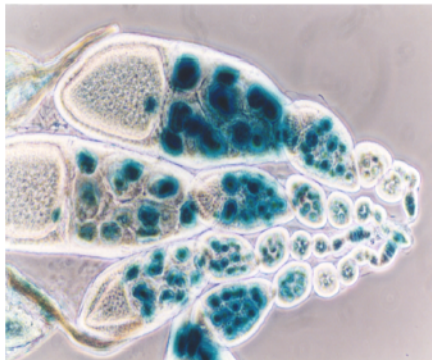
Le TSE est un phénomène qui a pu être observé dans les ovaires de drosophiles. Après une coloration au X-gal, on peut voir de la répression dans les noyaux des cellules nourricières et non dans les cellules folliculaires somatiques entourant la chambre à œuf. D'autre part, des transgènes cibles s'exprimant dans différents tissus (glande salivaire, cerveau larvaire, corps

gras) et disques imaginaux larvaires (ailes, pattes et yeux) ont été testés pour leur capacité à réprimer une cible. Dans aucun de ces tissus de la répression n'a pu être observée. L'utilisation de transgènes cibles s'exprimant dans les testicules ne montre pas de répression dans ce tissu non plus. Cette répression est donc spécifique de la lignée germinale femelle (Josse *et al.* 2008).

L'effet maternel

Tout comme la répression de l'élément *P*, le TSE présente un effet maternel. Lorsque le transgène silencer est apporté par la mère on observe une forte répression du transgène cible (80 à 100%) alors que lorsqu'il est apporté paternellement la répression du transgène cible est plus faible dans la descendance (0 à 40%) (Figure 11). L'effet maternel montre bien qu'il existe une composante cytoplasmique associée au TSE. La femelle pourrait déposer cette composante dans l'embryon ce qui permettrait ensuite l'établissement de la répression.

♀ *P-lacZ cible* x ♂ *P-lacZ silencer*



♀ *P-lacZ silencer* x ♂ *P-lacZ cible*

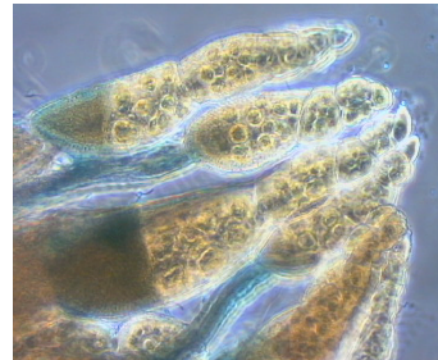


Figure 11. Observation de l'effet maternel de descendants issus des croisements indiqués pour chaque cas. A droite, on observe de la répression dans les cellules nourricières qui sont « blanches » lorsque le transgène silencer est apporté par la mère alors qu'à gauche on peut distinguer une coloration bleue quand le transgène silencer est transmis paternellement, l'ovaire ne montrant aucune répression (Josse *et al.* 2007).

Par ailleurs des expériences reposant sur l'effet maternel ont permis de mettre en évidence que les capacités de répression d'un transgène transmis paternellement s'acquièrent au fil des générations (environ 6) jusqu'à atteindre un niveau de TSE identique à un transgène transmis maternellement (Figure 12).

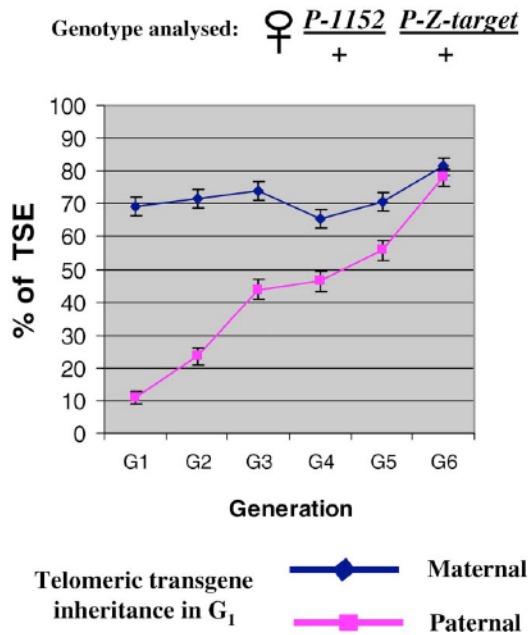


Figure 12. Dynamique d'acquisition des propriétés répressives pour un transgène silencer (*P-1152*) transmis maternellement (bleu) et paternellement (rose). La cible utilisée ici (*P-Z-target*) est *Pco1*. Le pourcentage de TSE est le niveau de répression du transgène (Josse *et al.* 2007).

Variegation

Lorsqu'on croise une femelle possédant un transgène silencer avec un mâle possédant un transgène cible, la répression de la cible peut parfois être incomplète. Ceci se caractérise par une distribution aléatoire de l'expression de LacZ de type ON (chambre à œuf bleue), ou OFF, (chambre à œuf réprimée et donc blanche), parmi les chambres à œuf au sein d'une même ovariole (Josse *et al.* 2007) (Figure 13). Ce phénomène est appelé variegation.

A l'heure actuelle, nous ne connaissons pas clairement les mécanismes sous-jacents liés à la variegation, on suppose qu'elle est associée à un état chromatinien du transgène cible différent selon les chambres à œufs.



Figure 13. Illustration du phénomène de la variegation dans les ovaires issus de la descendance d'un croisement d'une souche portant un silencer *P-LacZ* télomérique (*P-1152*) par une cible *P-lacZ* euchromatique (*BC69*) (Josse *et al.* 2007). On peut observer des ovarioles complètement réprimées (E) ou bien totalement bleues où le transgène s'exprime (B). De la variegation est visible parmi certaines ovarioles présentant à la fois des chambres à œufs où le gène *lacZ* est réprimé (blanc) ou exprimé (bleu) (A, C et D).

Indépendant des séquences de l'élément *P*

Jusqu'à présent, lorsqu'on parle de TSE, ce sont des transgènes silencer et cible *P-lacZ* qui sont utilisés. Or on pourrait imaginer que toutes les propriétés associées au TSE soient dues à la nature même du transgène grâce à la présence de séquences de *P* dans le silencer ainsi que dans la cible. Afin de montrer que le TSE est indépendant des séquences de *P*, des transgènes cibles *piggyBac-otu-lacZ* (*PBoL*) ont été générés. Ces transgènes ne possèdent pas de séquences de *P* et la seule région homologue est donc le gène *lacZ* qui est sous le contrôle d'un promoteur spécifique de la lignée germinale (*otu*). Quand on croise une femelle portant un transgène silencer télomérique avec un mâle portant un transgène cible *PBoL*, on observe de la répression, un effet maternel et de la variegation. On peut donc dire que le TSE ne dépend pas des séquences de l'élément *P* et peut être mise en place par une homologie de séquence correspondant seulement au gène *lacZ* (Poyhonen *et al.* 2012) (Figure 14). Ceci confère à ce mécanisme un caractère général.

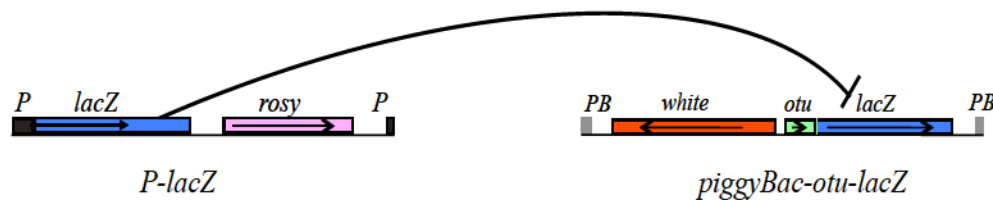


Figure 14. Répression d'un transgène cible *PBoL* par un transgène *P-lacZ*. Une femelle portant un transgène silencer *P-lacZ* est capable de réprimer en *trans* un transgène cible *piggyBac-otu-lacZ* indépendamment de la présence de séquences de *P*. Cette répression peut avoir lieu seulement grâce à l'homologie de séquence entre les séquences de *lacZ*.

D.4 Les acteurs moléculaires du TSE

Le TSE est un mécanisme complexe soumis à différentes propriétés montrant l'existence d'une composante chromosomique ainsi que d'une composante cytoplasmique. En effet, on observe du TSE que lorsque le transgène silencer est inséré dans les TAS uniquement lorsqu'il est transmis maternellement. De plus, le TSE peut-être incomplet et présenter de la variegation suggérant un mécanisme dépendant de la structure de la chromatine pour la mise en place de la répression. Au laboratoire, un crible de mutants de gènes candidats a donc été mené parmi les gènes impliqués dans la formation de l'hétérochromatine et dans la biogenèse des petits ARN.

Il a été montré que la répression par les transgènes *P-lacZ* télomériques est affectée en contexte mutant pour le gène *Su(var)205* codant la protéine HP1, composante de

l'hétérochromatine, et le gène *Su(var)3-7* qui code pour une protéine partenaire d'HP1 composante de l'hétérochromatine télomérique et centromérique. D'autre part, les gènes *Aubergine (Aub)*, *armitage*, *homeless* et *piwi* impliqués dans la biogenèse des piRNAs affectent aussi la répression par les transgènes *P-lacZ* (Josse *et al.* 2007).

Un peu plus tard, des mutants pour les gènes *squash* et *zucchini (zuc)* qui sont également impliqués dans la biogenèse des piRNA ont été testés sur le TSE (Todeschini *et al.* 2010). Le TSE est affecté par ces mutations et une perte des petits ARN compris entre 20 et 30nt a pu être observée dans ces mutants grâce à des expériences de RNase protection. Plus récemment, avec l'arrivée des analyses de données de séquençage à haut débit des petits ARNs ovariens, l'existence de piRNA homologues aux transgènes silenciers télomériques a pu être montrée. Leur présence requiert l'héritage maternel apporté par le transgène silencier (de Vanssay *et al.* 2012). Le TSE est donc un outil que l'on peut utiliser pour l'étude de la répression des éléments transposables en lignée germinale. Toutes ces données nous montrent l'implication à la fois de l'hétérochromatine et des petits ARN (piRNA) dans la répression des transgènes *P-lacZ* insérés dans les TAS.

Avant de vous décrire la biogenèse de ces piRNA, je m'attacherai d'abord à vous présenter dans le chapitre suivant le site 1A et plus précisément les TAS de *Drosophila melanogaster* responsables de la mise en place de cette répression et qui sont au cœur de ce projet de thèse.

Chapitre 2 : Organisation des régions télomériques chez *D. melanogaster*

Les télomères sont des structures répétées situées aux extrémités des chromosomes linéaires. Leur rôle est d'assurer la stabilité du génome évitant le raccourcissement des chromosomes au cours des générations. Bien que la fonction des télomères soit de protéger les chromosomes, il peut exister différents types de structures télomériques. Chez la plupart des eucaryotes, les télomères sont constitués de courtes répétitions d'une séquence nucléotidique (autour de 6pb) maintenues grâce à l'activité de la télomérase. Cependant, on ne retrouve pas cette enzyme chez les diptères (Sasaki and Fujiwara 2000). Chez la drosophile, des répétitions en tandem de trois rétrotransposons sans LTR spécifiques, du télomère *HeT-A*, *TAHRE* et *TART* (*HTT*), compensent la perte de la télomérase. Ainsi la longueur du télomère est maintenue par transpositions successives de ces ET aux extrémités chromosomiques. Dans ce chapitre, je vous décrirai toute la région télomérique de *Drosophila melanogaster* incluant la CAP, les *HTT* et les régions sub-télomériques appelées TAS auxquelles je me suis intéressée plus particulièrement durant cette thèse.

A. Les TAS

Les TAS pour « Telomeric Associated Sequences » sont des régions souvent référencées comme étant télomériques mais sont en réalité en position sub-télomériques. Elles se situent entre les télomères, strictement parlant *HTT*, et les premiers gènes euchromatiques (Figure 15). Ce sont des séquences hétérochromatiques répétées en tandem. Bien que ces régions existent chez beaucoup d'organismes, leurs séquences et leurs structures diffèrent considérablement d'une espèce à l'autre. D'autre part, même au sein d'une cellule, la structure des TAS peut-être différente d'un télomère à l'autre. Chez *Drosophila melanogaster*, il existe deux types de TAS : les TAS-R et les TAS-L (Mason, 2008).

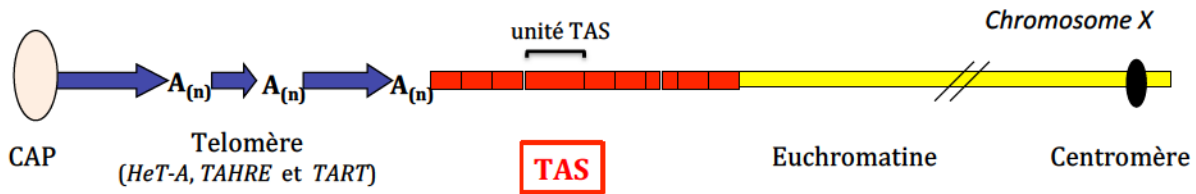


Figure 15. Détail d'une région télomérique. Cette région se compose de trois parties : en jaune, la CAP, en bleu, les télomères composés des ET *HeT-A*, *TAHRE* et *TART* et en rouge, on retrouve le locus TAS composé de plusieurs unités de tailles variables avant les premiers gènes.

A.1 Structure des TAS

A.1.1 La famille des TAS-R

Sous le terme « famille des TAS-R » sont regroupés les séquences TAS qui sont présentes sur les bras droits des chromosomes 2 et 3 (*TAS-2R* et *TAS-3R*) ainsi que sur le bras gauche du chromosome *X* (*TAS-X*) (Karpen and Spradling 1992). Ces TAS ont une structure particulière. En effet, chaque locus se compose de plusieurs répétitions que j'appellerai « unité TAS » (Figure 15). Ainsi selon le nombre d'unités, la taille d'un locus TAS peut varier d'un télomère à l'autre. Lorsque j'ai commencé cette étude, la structure d'une sous-unité de la famille des TAS-R était la suivante : trois répétitions de 170pb d'une partie du LTR d'un rétrotransposon appelé *Invader4* et une séquence ne correspondant à aucun ET connu (Figure 16). Il y a 9 insertions décrites de cet élément dans l'euchromatine, dont deux copies complètes, dans le génome de *D. melanogaster* séquencé (Flybase.org).

Le locus correspondant au *TAS-X* est le plus grand et le mieux caractérisé. Il a été séquencé à partir d'un mini-chromosome artificiel (Karpen and Spradling 1992). Le clone séquencé se compose dans l'ordre d'un fragment télomérique d'un *HeT-A*, de quatre sous-unités de TAS appelées TAS A, TAS B, TAS C et TAS D qui sont suivies de deux répétitions de 0.9kb appelées 0.9A et 0.9B. Parmi ces quatre répétitions, le TAS A est la plus petite. Elle fait 205pb et est le reflet d'une cassure chromosomique réparée par un *HeT-A* que l'on retrouve en 5' de cette séquence. Les TAS B et C font respectivement 1872 et 1857pb et sont identiques à plus de 95%. Enfin le TAS D fait 1574pb, il possède une délétion de 313pb en 3'. Le TAS B étant le plus grand, il est utilisé comme séquence TAS de référence pour le chromosome *X*.

Les *TAS-2R* et *3R* ont respectivement une longueur de 1002 et 982pb (Levis *et al.* 1993), (Yin and Lin 2007) et sont identiques à plus de 98% (Abad *et al.* 2004b). Ils ont en commun avec

le *TAS-X*, les trois répétitions du LTR d'*invader 4*. Le *TAS-2R* possède une répétition d'*Invader 4* en plus par rapport aux deux autres TAS-R (Figure 16). De plus, grâce à des expériences d'hybridation *in situ* sur chromosomes polytènes de larves de drosophile (voir matériels et méthodes) à l'aide d'une sonde faite à partir de la séquence *TAS-X*, on peut révéler les trois télomères 2R, 3R et XL. Ce qui peut être corrélé à leur degré d'identité (Karpen and Spradling 1992), (Andreyeva *et al.* 2005).

Dans la souche de drosophile séquencée, qui est la souche de référence $y^1; cn^1 bw^1 sp^1$ appelée « *Iso-1* », on peut dénombrer six répétitions complètes du *TAS-3R*, une répétition complète du *TAS-2R* mais on ne retrouve pas le *TAS-X*. En effet, les séquences TAS sont connues pour être polymorphiques et le *TAS-X* est celui dont l'absence a été le plus fréquemment observée dans les laboratoires (Levis RW, communication personnelle). Ces régions sont donc des régions assez complexes à cause de leur structure répétée et de leur polymorphisme.

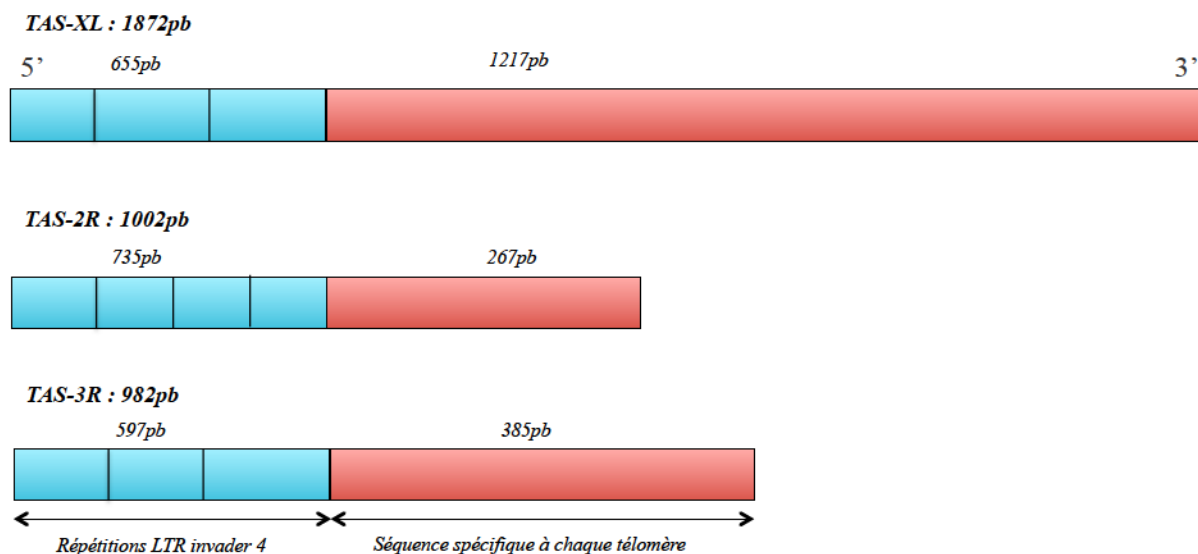


Figure 16. Détail des sous-unités *TAS-X*, *TAS-2R* et *TAS-3R*. En bleu, on retrouve les répétitions du LTR d'*Invader 4* et en rouge les séquences qui sont spécifiques à chacun des trois TAS des trois télomères. Pour le *TAS-X*, c'est le TAS B qui est représenté.

A.1.2 La famille des TAS-L

La famille des TAS-L regroupe les TAS présents sur les bras gauches des chromosomes 2 et 3 appelés *TAS-2L* et *TAS-3L*. Contrairement à la structure des TAS-R, celle des TAS-L est plus simple. En effet, un locus TAS-L se compose d'un mini-satellite de 461pb répété plusieurs fois (Walter *et al.* 1995). On ne retrouve aucune séquence d'ET ni la séquence du LTR d'*Invader 4* dans la composition de ces TAS. Les deux locus *TAS-2L* et *TAS-3L* sont identiques à 99,5%, ils ne diffèrent que par la présence d'un nucléotide (Abad *et al.* 2004b). D'autre part, dans *Iso-1* on dénombre 11 et 43 répétitions complètes du mini-satellite de 461pb pour les *TAS-2L* et *TAS-3L* respectivement. Des expériences d'hybridation *in situ* sur chromosomes polytènes en utilisant une sonde faite à partir de la séquence du *TAS-2L* ont permis de mettre en évidence cette séquence sur le télomère du chromosome 2L ainsi que le 3L. Cependant, aucun marquage n'a pu être observé sur le chromosome X (Walter *et al.* 1995). La famille des TAS-R et des TAS-L ne semble donc pas partager de régions homologues, détectables par cette technique. Récemment, grâce à l'alignement de séquences entre le *TAS-X*, les répétitions de 0.9kb et les TAS-L on retrouve une séquence commune de 160pb identique à 77% parmi ces trois structures sub-télomériques (Mason and Villasante 2014).

Ces données sont les seules disponibles à l'heure actuelle pour ces locus. Ceci malgré le rôle central joué par le *TAS-X* dans la mise en place de la répression de l'élément *P* et du TSE. Un des aspects de cette thèse a justement consisté à redéfinir les structures et à caractériser de manière plus fine les TAS-R et les TAS-L.

A.1.3 Le cas du chromosome 4 et du chromosome Y

Par rapport aux autres chromosomes de *Drosophila melanogaster* dont la taille varie de 23 à 32Mb, les chromosomes 4 et Y sont très petits. Leur taille est respectivement de 1.35 et 3.67Mb (Berkeley Drosophila Genome project). Suite à l'assemblage du télomère du chromosome 4 aucune structure similaire aux TAS-R ou aux TAS-L n'a pu être annotée. A la place, on retrouve entre les *HTT* et le dernier gène, une « zone de transition » de 5.4kb qui est un mélange de fragments d'ET télomériques (*HeT-A* et *TART*) ainsi que de fragments d'ET non télomériques (*CR1a*, *1360*, *ninja*) (George *et al.* 2006). Le chromosome Y est complètement hétérochromatique rendant son assemblage très compliqué. Jusqu'à présent aucune séquence présentant les propriétés des séquences TAS n'a pu être trouvée.

Une hypothèse est que l'hétérochromatine du chromosome *Y* serait différente de celle des régions télomériques des autosomes (Pardue and DeBaryshe 2008, Wang *et al.* 2014).

A.2 Les protéines associées aux TAS

Un outil appelé PICH (Proteomic analysis of Isolated Chromatin segments) a été développé afin d'identifier les protéines interagissant physiquement avec un locus spécifique (Antao *et al.* 2012). Pour valider leur méthode les auteurs ont utilisé les séquences TAS comme séquence cible. La chromatine a été isolée à partir de cellules embryonnaires somatiques (Kc) et les protéines sont identifiées par spectrométrie de masse. Un certain nombre de protéines interagissant avec la famille des TAS-L et des TAS-R ont ainsi pu être identifiées (Table 1 et Table 2).

Rank	Name	Domain or function
1	XNP	Chromatin remodeler
2	Rrp1	AP endonuclease
3	Dip3	Myb/SANT-like domain, BESS motif
4	CG8289	Chromodomain
5	Bj1	RCC1 superfamily
6	borr	Chromosome passenger complex
7	ial	Chromosome passenger complex
8	Trl	GAGA factor, BTB/POZ domain
9	CG8290	
10	CG3163	myb/SANT-like domain
11	smt3	SUMO
12	crol	Zn finger
13	CG4004	Myb/SANT-like domain
14	Orc2	Origin recognition complex
15	CG33691	
16	D1	AT hook like
17	Mi-2	Chromatin remodeler
18	Caf1	WD40 repeats, histone binding
19	GAG	GAG protein of Gypsy element
20	Mcm2	Minichromosome maintenance complex
21	CG1240	SWIB/MDM2, Dek domains
22	mu2	
23	psq	BTB/POZ domain
24	Su(var)2-10	Zn fingers, SAP and PINIT domains
25	LBR	ICMT domain
26	fbl6	F-box domain, leucine-rich repeats
27	Su(var)3-9	Chromodomain, SET domain
28	ran	GTPase
29	CG1910	

Table 1. Protéines associées aux TAS-L triées par ordre de confiance (Antao et al, 2012)

Rank	Name	Domains/function
1	Rrp1	AP endonuclease
2	Bj1	RCC1 superfamily
3	pont	AAA+ ATPase
4	Trl	GAGA factor, BTB/POZ domain
5	pita	Zn finger, C2H2 type
6	CG8289	Chromodomain
7	Mi-2	Chromatin remodeler
8	crol	Zn finger
9	D1	AT hook like
10	XNP	Chromatin remodeler
11	LBR	ICMT domain
12	pzg	
13	CG7946	PWWP domain, LEDGF domain
14	mor	SANT, SWIRM, RSC8 domains
15	Rfc38	AAA+ ATPase, RFC small subunit
16	Kdm2	H3K4 demethylase
17	CG1240	SWIB/MDM2, Dek domains
18	Dref	BED Zn finger
19	Chro	Chromodomain
20	HP1c	Chromodomain, chromo shadow
21	gp210	
22	smt3	SUMO
23	crp	Helix-loop-helix domain
24	borr	Chromosome passenger complex
25	Ubqn	Ubiquitin-like domain
26	ran	GTPase
27	Klp3A	Kinesin motor domain
28	BEAF-32	BESS motif, BED Zn finger
29	Caf1	WD40 repeats, histone-binding
30	Cp190	BTB/POZ domain
31	Cpr	NADPH-dependent flavin mononucleotide reductase
32	Parp	PARP, WGR, Zn finger, BRCT domains
33	Adf1	Myb/SANT-like domain
34	osa	ARID/BRIGHT DNA binding domain
35	Dek	SAP domain
36	Orc1	Origin recognition complex
37	HP1b	Chromodomain, chromo shadow
38	Rpb5	RNA polymerase II subunit
39	bocksbeutel	LEM domain
40	Nipped-B	Adherin
41	ial	Chromosome passenger complex
42	CG17078	
43	GAG	GAG protein of Gypsy element
44	CG1910	
45	ham	Zn finger
46	Dip3	Myb/SANT-like domain, BESS motif
47	wds	WD40 domain
48	zf30C	
49	Snr1	SNF5 superfamily

Table 2. Protéines associées aux TAS-R triées par ordre de confiance (Antao et al, 2012)

On retrouve certains gènes comme HP1 qui ont déjà été décrits comme interagissant avec les TAS (Frydrychova *et al.* 2008). Une partie des gènes candidats trouvés par cette technique ont été validés par des expériences de ChIP-qPCR. L'analyse par « gene ontology » fait ressortir les termes suivants : régulation de l'expression de gènes, fixation à la chromatine et fixation à l'ADN. Ces fonctions semblent bien correspondre à ce type de régions. Pour cette analyse les auteurs ont utilisé une séquence correspondant à la séquence d'*Invader4* pour leur expérience de PICCh. Peut-être que toutes les protéines qu'ils trouvent ne sont pas présentes sur le locus TAS-R car il y a deux copies d'*Invader 4* complètes dans le génome.

Les TAS possèdent certains motifs pouvant être reconnus par la famille des protéines Polycomb (PcG) (Boivin *et al.* 2003). Ces protéines permettent de maintenir un état répressif des gènes en compactant la chromatine. La marque répressive H3K27me3 a été montrée comme étant présente sur les TAS par des expériences de FISH sur chromosomes polytènes. Celle-ci pourrait être déposée par E(z), un membre de la famille des Polycomb (Andreyeva *et al.* 2005). Ces résultats suggèrent que les TAS possèdent une chromatine répressive.

A.3 Le rôle des TAS

L'implication des TAS a été décrite dans plusieurs mécanismes différents sans pour autant permettre de pouvoir leur attribuer un rôle défini dans le génome, en voici quelques exemples. Tout d'abord, le *TAS-X* a été décrit comme étant un point chaud d'insertion de l'élément *P*. En effet, un élément *P* inséré dans le *TAS-X* est capable de réprimer l'expression de dizaines de copies de *P* dispersées ailleurs dans le génome (Ronsseray *et al.* 1996). Parallèlement, les TAS sont aussi impliqués dans le mécanisme de TSE car les transgènes *P-LacZ* deviennent silencieux lorsqu'ils sont insérés dans les TAS-R (Josse *et al.* 2007, Josse *et al.* 2008). Ces deux observations nous indiquent que les TAS possèdent des propriétés spéciales, un ET ou un transgène inséré dedans semble s'approprier les capacités associées à ce locus hôte dans les cellules germinales.

Un autre rôle des TAS a aussi été décrit dans les tissus somatiques. Des transgènes *P* portant un gène rapporteur *white* ont été très largement étudiés ces 20 dernières années. Lorsque ces transgènes s'insèrent dans l'euchromatine, on peut observer une pigmentation rouge de l'œil de drosophile révélant une forte expression de *white*. En revanche quand le même transgène est dans une région hétérochromatique, l'œil présente une pigmentation mosaïque (rouge et blanche). Ce phénomène est appelé « position effect variegation » ou PEV et quand le transgène est dans le télomère on dit alors qu'il s'agit de « telomeric position effect » ou TPE.

(Mason *et al.* 2008). Le TPE concerne aussi bien les TAS-L que les TAS-R. Un dernier rôle parfois évoqué dans la littérature mais très peu décrit serait le rôle des TAS dans le positionnement des chromosomes à l'intérieur du noyau.

B. Les rétrotransposons télomériques : *HeT-A*, *TAHRE* et *TART*

Chez la plupart des organismes eucaryotes, les télomères sont composés de courtes répétitions nucléotidiques mais chez la drosophile ce sont trois rétrotransposons sans LTR appelés *HeT-A*, *TAHRE* et *TART* qui jouent le rôle du télomère. Il s'agit ici d'un cas de domestication des ET permettant de maintenir l'intégrité de la région télomérique.

B1. *HeT-A*, *TART* et *TAHRE*

HeT-A, *TART* et *TAHRE* (ou *HTT*) sont des rétrotransposons de type LINE ne possédant pas de LTR spécifiques des télomères qui sont le résultat de leur transposition aux extrémités des chromosomes. Parmi ces ET, c'est *HeT-A* qui est le plus abondant dans le génome. C'est un ET télomérique dont on peut trouver quelques fragments dans l'hétérochromatine péricentromérique (Traverse and Pardue 1988). Contrairement aux rétrotransposons sans LTR canoniques, *HeT-A* possède une seule ORF (ORF1) codant pour une protéine GAG-like. Il lui manque donc l'ORF 2 codant pour la reverse transcriptase (RT) et l'endonuclease. L'analyse d'un fragment terminal d'ADN dans les années 90 a permis la découverte du second ET télomérique *TART* (telomere-associated retrotransposon) (Levis *et al.* 1993). C'est un rétrotransposon sans LTR classiques avec deux ORF (ORF 1 et ORF2). Enfin, il y a *TAHRE* (telomere-associated and *Het-A* related element) qui est l'élément le moins abondant (Abad *et al.* 2004a). Tout comme *TART*, cet élément possède aussi deux ORF (Figure 17). Une hypothèse est que ces trois ET télomériques auraient une origine commune. En effet, les régions 5'UTR, 3'UTR et l'ORF1 de *HeT-A* présentent des homologies de séquences avec *TAHRE*. D'autre part, des comparaisons entre les ORF 1 et 2 de *TART* avec celles de *TAHRE* semblent aussi indiquer des similitudes entre ces deux ET. Une dernière hypothèse est que *HeT-A* aurait dérivé de *TAHRE* suite à la perte de l'ORF2 (Abad *et al.* 2004a). En ce qui concerne les promoteurs de ces trois ET, ceux d'*HeT-A* et *TAHRE* seraient dans leur régions 3'UTR alors que *TART* en possèderaient deux en 3' et 5'UTR (Zhang and Rong 2012). Comment ces ET permettent-ils l'élongation du télomère ?

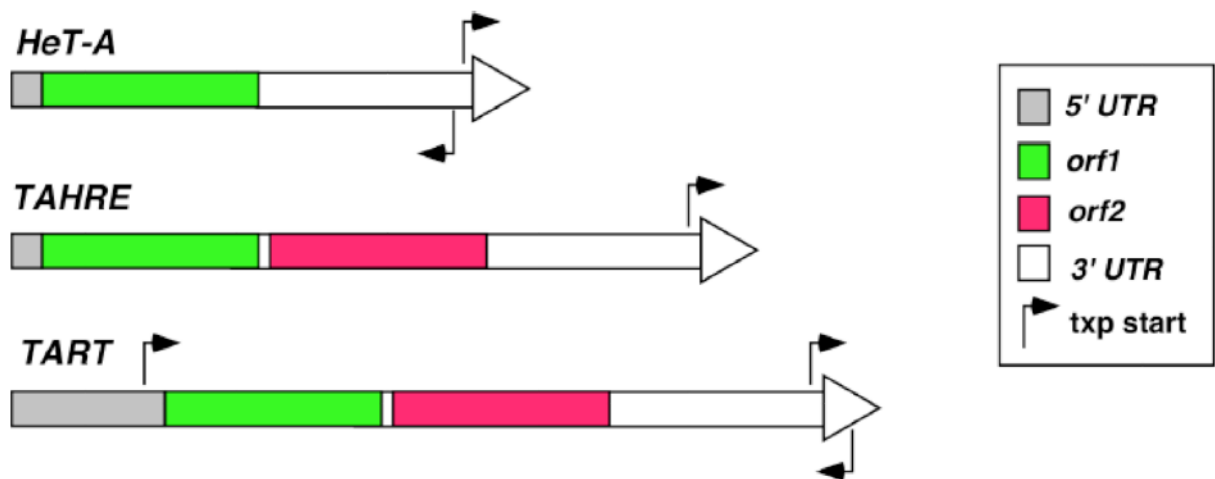


Figure 17. Représentation schématique des trois ET télomériques : *HeT-A*, *TAHRE* et *TART* (*HTT*) chez *Drosophila melanogaster*. Les flèches représentent les positions approximatives des sites d'initiation de la transcription (txp start) (Zhang and Rong 2012).

B2. Mécanisme de transposition et élongation du télomère

La maintenance du télomère chez la drosophile dépend de la transposition en tandem des *HTT* aux extrémités des chromosomes. Un modèle permettant d'expliquer l'élongation du télomère a été décrit. Les *HTT* sont tout d'abord transcrits puis exportés du noyau vers le cytoplasme où sont traduites les protéines GAG-like et la reverse transcriptase. *HeT-A* n'ayant pas d'ORF2 codant la reverse transcriptase, celle-ci pourrait lui être fournie *in trans* par *TART* ou *TAHRE*. Une partie de ces transcrits est ensuite prise en charge par ces protéines qui viennent d'être traduites afin de pouvoir retourner dans le noyau où ils sont dirigés spécifiquement vers les extrémités du chromosome où a lieu la transposition (pour revue (Zhang and Rong 2012);(Fulcher *et al.* 2014)) (Figure 18). Chaque événement de transposition permet d'augmenter la taille du télomère. Cependant on ne sait pas comment ces transcrits sont ciblés aux extrémités et quel est le mécanisme de transposition de ces ET ? Une hypothèse est qu'il y aurait une interaction possible entre les protéines constituant la cap et la protéine GAG, où la machinerie de transposition pourrait être recrutée (Zhang and Rong 2012).

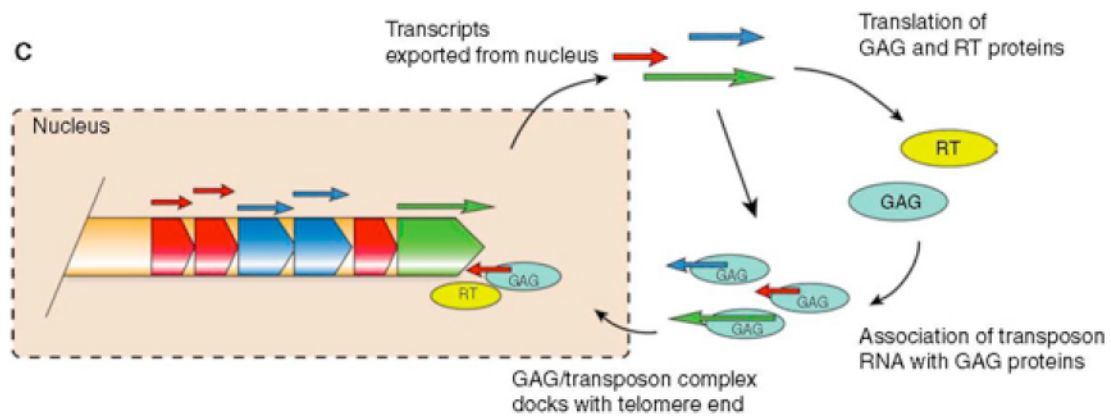


Figure 18. Mécanisme de transposition des *HTT* aux télomères chez *Drosophila melanogaster*. Les *HTT* sont représentés en rouge, bleu et vert. Les transcrits sont exportés vers le noyau (flèches rouges, bleues et vertes) où ils sont traduits en GAG et reverse transcriptase (RT) permettant ainsi leur retour dans le noyau aux extrémités chromosomiques (Fulcher *et al.* 2014).

C. La cap

Il existe une structure appelée la « cap » (casquette en anglais) aux extrémités des chromosomes qui assure la protection des télomères contre la fusion télomérique et les systèmes de réparation de l'ADN. Deux groupes de protéines composent la cap, il y a le complexe de la terminin chez la drosophile qui forme la structure même de la cap et d'autres protéines pouvant interagir directement ou indirectement avec ce complexe. Au total treize protéines formant la cap ont pu être identifiées chez la drosophile car leur mutation provoque un phénotype de fusion télomérique dans les cellules mitotiques du cerveau larvaire (Table 3). La terminin est le complexe formé par les protéines HOAP (HP1/ORC Associated Protein), Moira (Moi), verrocchio (ver) et HipHop (HP1 interacting protein). Ces protéines ont pu être localisées aux télomères par des expériences d'immunomarquages sur chromosomes polytènes (Raffa *et al.* 2011).

Ce complexe est spécifiquement recruté au télomère et est homologue fonctionnellement à la shelterin (le complexe de capping chez l'homme) (Figure 19). Il existe aussi d'autres protéines présentes au niveau du télomère mais celles-ci ne font pas partie de la terminin car elles ne sont pas spécifiquement localisées aux télomères. On retrouve les protéines : ATM, HP1a, Mre11, Nbs, Rad50 UbcD1 et Woc (Raffa *et al.* 2013). Cette dernière protéine a un statut important car elle code pour un facteur de transcription. Elle a été localisée sur chromosomes polytènes au niveau télomérique ainsi que dans certaines régions euchromatiques (Font-Burgada *et al.* 2008). Woc pourrait être un facteur de transcription de la région télomérique et aussi des TAS. En effet, cette protéine a été montrée être en complexe

avec HP1c et Row (Related of woc) au niveau des TAS-R dans les cellules Kc, ceci à été montré par des expériences en western blot (Antao et al, 2012). Par ailleurs, ces trois protéines ont aussi pu être co-localisées au niveau de plusieurs bandes sur chromosomes polytènes (Font-Burgada *et al.* 2008). Cependant la co-localisation avec une sonde spécifique des TAS et woc n'a pas encore été effectuée.

Gene name	Protein name	TFs/cell in mutants	Function outside telomeres	Human homolog	Function at human telomere
<i>cav</i>	HOAP	5	None known	None	–
<i>hiphop</i>	HipHop	Many (1)	None known	None	–
<i>moi</i>	Moi	5	None known	None	–
<i>ver</i>	Ver	5	None known	STN1 (3)	–
<i>Su(var)205</i>	HP1 α	4	Heterochromatin regulation; transcription factor	CBX5/ HP1 α	Yes
<i>eff</i>	UbcD1	0.7	E2 ubiquitin-conjugating enzyme	UBE2D2/; UbcH5b	Nd
<i>woc</i>	Woc	2	Transcription factor	ZMYM3/; ZNF261	Nd
<i>mre11</i>	Mre11	0.5	DNA repair	MRE11	Yes
<i>rad50</i>	Rad50	0.5	DNA repair	RAD50	Yes
<i>nbs</i>	Nbs	0.4	DNA repair	NBS1	Yes
<i>tefu</i>	ATM	0.6	Kinase; DNA damage response	ATM	Yes
<i>mei-41</i>	ATR	None (2)	Kinase; DNA damage response	ATR	Yes
<i>mus-304</i>	ATRIP	None (2)	DNA helicase; DNA damage response	ATRIP	Yes

Table 3. Protéines impliquées dans la prévention de la fusion télomérique dans les cellules mitotiques (Raffa *et al.* 2013).

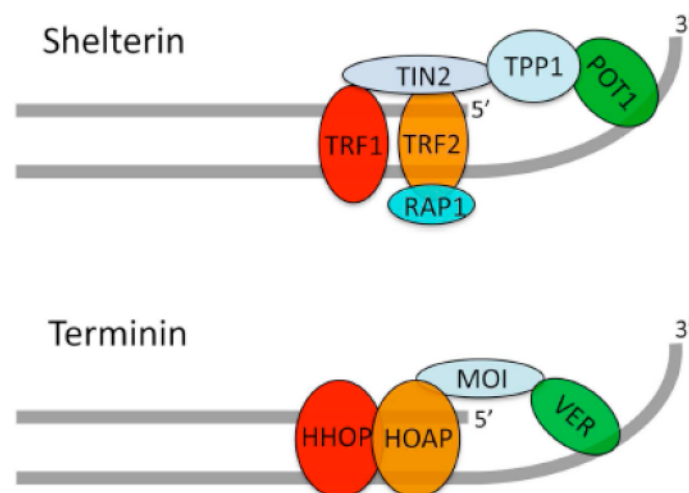


Figure 19. Modèle hypothétique de la structure de la terminin chez la drosophile et de la shelterin chez l'homme. HipHop (HHOP) et HOAP seraient présentes sur l'ADN double brin au niveau des télomères. La protéine Verrocchio (ver) est présente sur l'extrémité simple brin de l'ADN en 3'. Moira (Moi) n'interagit pas avec l'ADN mais servirait plutôt de connecteur pour le complexe HHOP/HOAP et ver. Ceci peut-être mis en parallèle avec le même complexe présent chez l'homme appelé Shelterin. Les mêmes couleurs sont utilisées par les protéines homologues chez la drosophile et l'homme (Raffa *et al.* 2013).

Les télomères de la drosophile sont donc une région complexe sous-divisée en trois parties, la cap, les *HTT* et les TAS. Au cours de ma thèse, je me suis intéressé plus particulièrement aux TAS et notamment à l'un de ses rôles qui n'avait pas été cité auparavant : celui de locus producteur de piRNA. En effet, ce sont les piRNA qui sont impliqués dans la répression des ET en lignée germinale. Dans le chapitre suivant, nous allons voir comment ces petits ARN sont produits et par quels mécanismes ils établissent la répression des ET.

Chapitre 3 : Les petits ARNs impliqués dans la régulation des ET en lignée germinale

A. Les PIWI-interacting RNA (piRNA)

Au laboratoire, nous étudions une classe particulière de petits ARN non codants : les piRNA. Ils sont associés aux protéines de la famille PIWI (Aubergine, Piwi et AGO3). Leur taille est comprise entre 23 et 29 nucléotides (Brennecke *et al.* 2007) ; (Gunawardane *et al.* 2007). Ces petits ARN sont produits spécifiquement dans la lignée germinale chez la drosophile par des locus distincts. Leur rôle est de réprimer l'expression des éléments transposables dans les cellules germinales assurant ainsi l'intégrité du génome de la future descendance.

J'exposerai dans cette partie les différents acteurs associés à la biogenèse de ces petits ARN ainsi que les mécanismes permettant la répression des éléments transposables connus à l'heure actuelle.

A.1 Les clusters producteurs de piRNA

Les piRNA dérivent de régions particulières du génome appelé « clusters de piRNA ». Ces clusters sont composés d'un catalogue de fragments d'éléments transposables (Figure 20). Chez la souris, 90% des piRNA seraient générés à partir d'une centaine de régions génomiques allant jusqu'à 100kb (Yamanaka *et al.* 2014) Chez la drosophile, 142 clusters producteurs de piRNA ont été identifiés et classés selon le nombre de piRNA qu'ils produisent (Brennecke *et al.* 2007). Ils représentent moins de 5% du génome et génèrent à eux seuls 90% des piRNA (Malone *et al.* 2009). Parmi ces clusters, on retrouve le *TAS-X*, *TAS-2R*, les *TAS-L* et *TAS-3R* qui sont respectivement les clusters n°3, 11, 79 et 97. Ces clusters sont situés majoritairement dans les régions péricentromériques et télomériques. Ceci soulève une question : comment les clusters sont-ils transcrits alors qu'ils se trouvent dans des régions recouvertes de marques hétérochromatiques répressives ?

Il existe deux types de clusters : les uni-strand clusters et les dual-strand clusters caractérisés par l'orientation des sens de transcription (Figure 20). Ainsi les uni-strand sont transcrits unidirectionnellement et sont exprimés majoritairement dans les cellules folliculaires somatiques, et les dual-strand clusters sont transcrits sur les deux brins de l'ADN et sont exprimés dans les cellules germinales ovariennes. Il existe cependant une exception, le cluster 2 qui est un

cluster germinale mais transcrit uni-directionnellement (Yamanaka *et al.* 2014); (Hirakata and Siomi 2016).

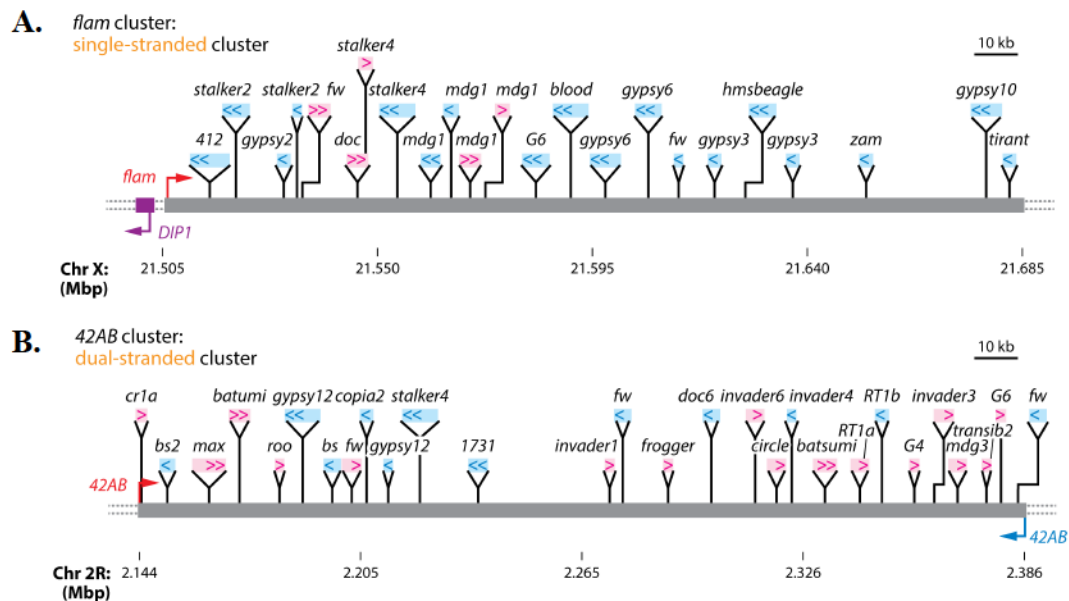


Figure 20. Représentation schématique de la structure des clusters de piRNA. A. *Flamenco*, un uni-strand cluster somatique. La flèche rouge indique le sens de transcription. **B.** *42AB*, un dual strand cluster. Les flèches rouge et bleue de part et d'autre du cluster représentent les deux sens de transcription. Les flèches bleues et roses indiquent le sens d'insertion des ET dans le cas A et B (Iwasaki *et al.* 2015).

A.2 La biogenèse des piRNA germinaux

Les piRNA germinaux sont produits majoritairement à partir de dual-strand clusters. Cette biogenèse a lieu dans les noyaux des cellules nourricières. Une des sources majeures de piRNA germinaux est le cluster *42AB*. Il se situe dans l'hétérochromatine péricentromérique du bras droit du chromosome 2. C'est aussi l'un des plus grands, il fait plus de 240kB. Il est constitué de plusieurs centaines de fragments d'éléments transposables tronqués insérés dans les deux orientations (Figure 20). Les piRNA produits à partir de ce cluster seront donc à la fois sens et antisens. Ce cluster produit plus de 30% des piRNA germinaux totaux (Brennecke *et al.* 2007).

Les dual-strands clusters semblent avoir une transcription particulière. Ils ne possèdent pas de promoteurs clairement identifiés. Ils sont transcrits par l'ARN polymérase II (ARN pol II) bien qu'ils soient recouverts d'une marque chromatinienne répressive H3K9me3. Les

transcrits issus de ces clusters n'ont pas de coiffe 5'methyl-guanosine, ni de terminateurs de transcription (Mohn *et al.* 2014), (Zhang *et al.* 2014). Ces transcrits dépourvus de coiffes devraient être dirigés vers la voie de la dégradation des ARN ce qui n'est pas le cas car ils permettent de générer des piRNA matures. Les dual-strands clusters ont donc une transcription que l'on peut qualifier de « non canonique ».

Plusieurs études ont apporté de nouveaux éléments permettant de comprendre le mécanisme de transcription des dual-strands clusters. Ainsi, il a été montré par des expériences de ChIP que Rhino, l'homologue de la protéine hétérochromatique 1 (HP1), est requise pour la production de piRNA (Klattenhoff *et al.* 2009). Rhino pourrait se lier à la marque H3K9me3 présente sur ces clusters via son chromodomaine. Cette marque étant déposée en amont par une histone méthyltransférase Eggless (Rangan *et al.* 2011). L'association de Rhino/H3K9me3 recrute ensuite la protéine Cutoff, homologue du facteur de terminaison de transcription Rai1 chez la levure (Kim *et al.* 2004), ainsi que la protéine Deadlock. Ce complexe formé par Rhino-Deadlock-Cutoff est appelé le complexe RDC et constitue à l'heure actuelle un facteur déterminant pour la définition des dual-strands clusters. La protéine Cutoff se lie et protège la partie 5' du transcrit contre le CAP binding complexe (CBC) résultant en l'absence d'épissage, de polyadénylation et de terminaison de transcription permettant une transcription continue des clusters (Mohn *et al.* 2014) (Figure 21). Le complexe RDC permettrait à la fois la transcription des clusters ainsi que leur protection contre la terminaison de transcription.

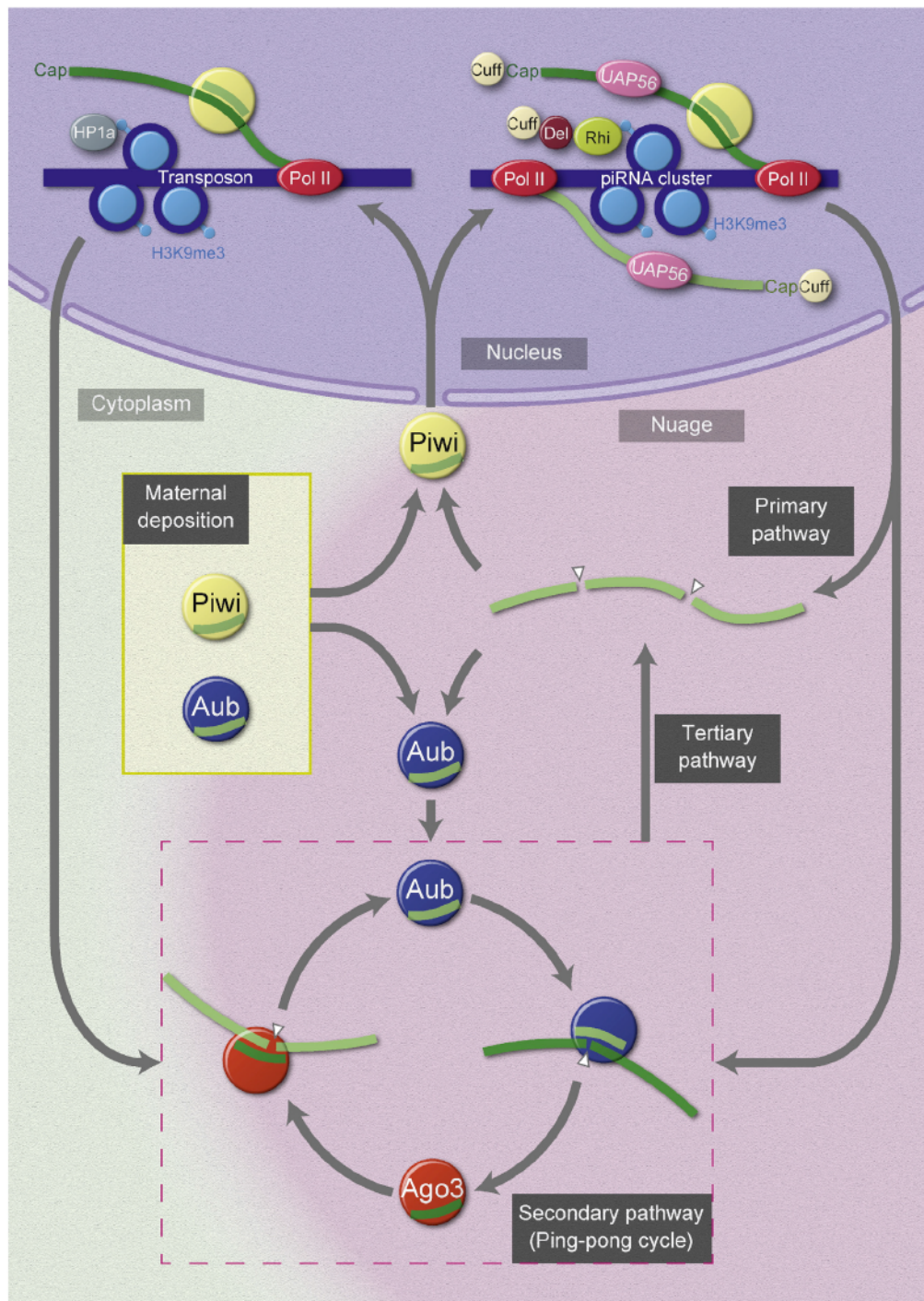


Figure 21. Vue d'ensemble de la biogenèse des piRNA germinaux. Dans le noyau, Rhino (Rhi) peut se fixer grâce à son chromodomaine à la marque H3K9me3. Ceci entraîne la fixation de Del (Deadlock) et Cuff (Cutoff) formant ainsi le complexe RDC. Cuff se fixe à l'extrémité 5' du transcrit naissant permettant la transcription non canonique des dual-strand clusters. La protéine UAP 56 va guider le transcrit jusqu'au nuage (Hirakata and Siomi 2016).

A.2.1 Les piRNA germinaux primaires

Les précurseurs des piRNA sont ensuite exportés vers le cytoplasme dans une structure dense aux électrons appelée le « nuage » (Eddy 1974) dont le composant principal, Vasa, est une hélicase possédant un domaine DEAD-box (motif fixation à l'ARN). On retrouve dans le nuage de nombreux facteurs associés à la biogenèse des piRNA (Pek *et al.* 2012). Un des acteurs connus pour ce passage vers le cytoplasme est l'hélicase nucléaire possédant aussi un domaine DEAD-box : UAP56 (Figure 21). Cette protéine co-localise en effet avec Rhino sous forme de foci dans le noyau près du pore nucléaire faisant face à Vasa dans le nuage de l'autre côté de l'enveloppe nucléaire (Zhang *et al.* 2014). De plus, les auteurs observent une perte du nuage ainsi qu'une baisse de piRNA associés aux dual-strand clusters en contexte mutant pour UAP56. Ils proposent ainsi un modèle où UAP56 accompagnerait les transcrits des clusters germinaux vers le nuage où ils seraient transmis à Vasa.

Une des caractéristiques des piRNA primaires matures est leur extrémité 5'-monophosphate avec une uridine en position 1 (1U) (Figure 22). Cependant on ne connaît pas les facteurs responsables de la formation de cette extrémité. De récentes études ont montré que la protéine Zucchini (Zuc) pourrait être un candidat idéal pour cette étape. En effet, cette protéine est capable de se fixer spécifiquement aux ARN simple brin sans spécificité de séquences. Contrairement à ce qui était décrit précédemment, en plus de son activité phospholipase D, Zuc possède aussi une activité d'endonucléase qui a été démontrée *in vitro* (Nishimasu *et al.* 2012). L'ARN clivé présente une extrémité 5'-monophosphate, une marque de reconnaissance des piRNA matures. Dans un ancien crible EMS (Schupbach and Wieschaus 1991), la protéine Zuc est associée à de la stérilité chez la femelle ainsi que d'autres protéines décrites plus tard comme intervenant dans la biogenèse des piRNA (Aub, Shutdown, Vasa, Squash, Cutoff et Deadlock). Plus récemment, il apparaît que lorsque Zuc est mutée la production de piRNA en lignée germinale est dérégulée. L'expression de deux rétrotransposons télomériques, *Het-A* et *TART*, de l'ET *roo* (Pane *et al.* 2007) ainsi que celle de transgènes *P-lacZ* (Todeschini *et al.* 2010) est affectée en contexte mutant pour Zuc. Les auteurs (Pane *et al.* 2007) observent par PCR quantitative en temps réel une augmentation d'expression d'un facteur 1000 et 15 pour *HeT-A* et *TART*, respectivement, suggérant que cette protéine est requise pour le silencing des ET en lignée germinale. Un dernier point à souligner dans cette étude est que les auteurs démontrent que la protéine Zuc serait présente dans le nuage et interagirait avec Aub.

Les piRNA dont l'extrémité a été définie par Zuc en 5' ne sont pas tout à fait encore matures car ils sont plus longs du côté 3'. La compréhension des processus moléculaires en jeu ainsi que la ou les protéines intervenant dans la dégradation de l'extrémité 3' constitue actuellement un axe majeur de recherche pour élucider la biogenèse des piRNA primaires germinaux. Deux cribles à grande échelle ont été menés à ce jour pour rechercher les protéines intervenant dans la biogenèse des piRNA, malheureusement la nucléase pouvant jouer ce rôle n'a pas encore pu être identifiée (Czech *et al.* 2013);(Muerdter *et al.* 2012).

Pour résumer, Le complexe RDC dirige la transcription non canonique des dual strand clusters. Le transcrit du cluster sort du noyau grâce à UAP56 et se retrouve dans le nuage où Zuc et une/ou plusieurs protéines inconnues effectuent les étapes de maturation de ces piRNA primaires avec 1U en 5'. Ces piRNA sont ensuite chargés sur les protéines Aub ou Piwi qui vont spécifier leur devenir. Les piRNA chargés par Piwi pourront effectuer de la répression transcriptionnelle (TGS « Transcriptional Gene Silencing ») car Piwi peut retourner dans le noyau. Si ces piRNA sont associés avec Aub dans le nuage, ils vont alors rentrer dans un cycle d'amplification appelé « ping-pong » et générer des piRNA dits secondaires pour effectuer de la répression post-transcriptionnelle (PTGS « Post-Transcriptional Gene Silencing »).

A.2.2 Les piRNA germinaux secondaires

Les piRNA primaires associés à Aub et Piwi possèdent un fort biais 1U et sont anti-sens aux ET (Brennecke *et al.* 2007) ; (Gunawardane *et al.* 2007). Dans le nuage, les complexes Aub-piRNA démarrent le cycle d'amplification des piRNA en résonance avec AGO3 grâce à leurs activités endonucléases pour produire des piRNA secondaires. Une autre source possible de piRNA chargés par Aub est le dépôt maternel dans l'embryon pouvant alimenter le ping-pong (Figure 21).

Les piRNA chargés par Aub sont anti-sens et ciblent les transcrits d'ET sens par complémentarité de séquence. Les protéines Argonautes clivent les ARN à la position 10 à partir de l'extrémité 5' de leur ARN guide. Les transcrits clivés par Aub auront donc un A en position 10 à cause de la complémentarité de séquence avec le 1U générant ainsi un piRNA sens. Ce piRNA est pris en charge par AGO3 qui à son tour pourra cliver un transcrit d'ET anti-sens générant donc un piRNA anti-sens qui sera repris en charge par Aub et ainsi de suite

(Figure 22). Le biais introduit par la présence d'un U en position 1 des piRNA primaires, et d'un A en position 10 des piRNA secondaires, produit des molécules ayant une complémentarité de 10pb et est à l'origine de ce que l'on appelle la « signature ping-pong ». Elle est le reflet de ce processus qui est une boucle entre Aub et AGO3, d'où le terme de « ping-pong ». Dans les dual-strand cluster, les ET étant insérés dans les deux sens, le ping-pong permet ainsi de produire un pool de piRNA secondaires à la fois sens et anti-sens dirigés contre les ET en lignée germinale. Les piRNA secondaires vont intervenir dans la régulation post-transcriptionnelle des ET. Récemment un nouveau pool de piRNA tertiaires a pu être décrit (Han *et al.* 2015); (Mohn *et al.* 2015), ils sont générés à partir des piRNA secondaires. Les extrémités 5' des piRNA sont produites par Aub et AGO3 mais on ne connaît pas le mécanisme pour dégrader les extrémités 3' pour former des piRNA matures secondaires. Un autre point à souligner est le statut particulier des piRNA chargés par Aub. En effet, une fois engagé dans le ping-pong les analyses bio-informatiques ne permettent plus de savoir si ces piRNA sont des piRNA primaires ou des piRNA secondaires.

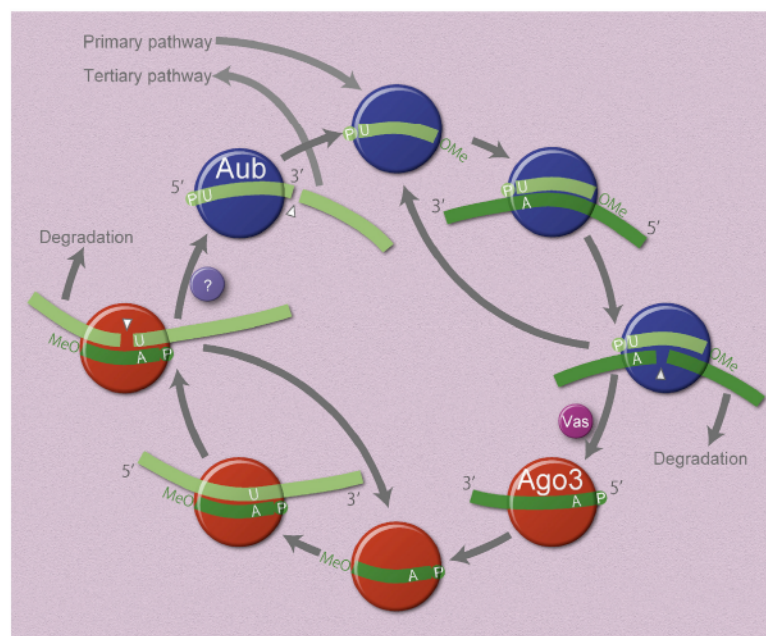


Figure 22. Modèle du ping pong chez la drosophile. Les complexes Aub-piRNA sont générés par la voie primaire (primary pathway) avec un U en position 1. Ces piRNA peuvent cibler par complémentarité de séquence des transcrits d'ET (en vert foncé) générant un A en position 10. Ce transcript alors pris en charge par AGO3 et peut cliver à nouveau par complémentarité de séquence un transcript issue d'un cluster de piRNA générant un U en position 1 qui peut alors être pris en charge par Aub. Ce processus va ainsi générer des piRNA secondaires. Les flèches blanches indiquent la position du clivage. D'autre part une partie des piRNA secondaires est aussi impliquée dans la production de piRNA tertiaires (Hirakata and Siomi 2016).

A.2.3 Les piRNA germinaux tertiaires

Une nouvelle population de piRNA appelés piRNA tertiaires ont pu être mis en évidence grâce à des analyses bio-informatiques (Han *et al.* 2015), (Mohn *et al.* 2015). Les piRNA secondaires seraient responsables de la production de piRNA tertiaires. Les piRNA produits par AGO3 générés lors du ping-pong sont appelés les piRNA « responder » et devraient normalement être pris en charge par Aub. Or ces « responder » piRNA participent à la production de piRNA appelés « trailer » qui sont associés avec Piwi. Les piRNA « responder » et « trailer » présentent une marque générée par Zuc qui se traduit par un intervalle de 27nt et un biais de 1U. Ceci permet d'augmenter la quantité de piRNA chargés par Piwi pouvant rentrer dans le noyau. Cette nouvelle population de piRNA pourrait cibler un répertoire plus large de séquences.

A.3 La biogenèse des piRNA folliculaires somatiques

Le cluster majeur produisant les piRNA somatiques est le cluster *Flamenco/COM* situé sur le chromosome *X*. Ce locus est notamment étudié pour la régulation des éléments *gypsy*, *Idefix* et *ZAM* (Pelisson *et al.* 1994), (Mével-Ninio *et al.* 2007), (Zanni *et al.* 2013). Les éléments transposables dans ce locus sont majoritairement insérés en anti-sens par rapport au sens de transcription. Ce cluster fait plus de 150kb et se comporterait comme une seule unité de transcription (Malone *et al.* 2009).

Les uni-strands clusters semblent être indépendants de Rhino car la perte de cette protéine ne semble qu'affecter les piRNA issus des dual-strands clusters (Mohn *et al.* 2014). De plus Rhino n'est pas exprimée dans les cellules somatiques folliculaires où seuls les uni-strands clusters sont actifs. Les mécanismes liés à la transcription des uni-strands cluster ne sont pas encore clairement connus. Contrairement à la biogenèse des piRNA germinaux, celle des piRNA somatiques est plus simple. Celle-ci a lieu dans les cellules folliculaires somatiques entourant la chambre à œuf. Les protéines Aub et AGO3 n'y sont pas exprimées : on retrouve seulement la protéine Piwi. De ce fait, il n'y a pas de processus similaire au ping-pong permettant l'amplification des piRNA dans les cellules somatiques.

Les précurseurs des piRNA sont d'abord transcrits à partir des uni-strands cluster par l'ARN polymérase II de façon canonique. Le facteur de transcription de *Flamenco* serait Ci (Cubitus interruptus). Les sites de fixation prédits pour Ci muté et ceci entraîne une diminution d'un

facteur 10 de l'expression de *Flamenco*. Par ailleurs, les auteurs montrent qu'il existe un épissage pour ce transcrit donnant lieu à différentes tailles de précurseurs de piRNA (Goriaux *et al.* 2014). Ces transcrits primaires sont ensuite exportés vers le cytoplasme où ils sont pris en charge par la protéine Zuc située sur la surface des mitochondries. Cette protéine intervient lors de l'étape de maturation des piRNA. Dans les cellules en culture OSC et OSS (lignées de cellules ovariennes somatiques), des mutations dans le domaine catalytique de Zuc entraînent une forte accumulation de précurseurs de transcrits ainsi que des piRNA intermédiaires dans le cytoplasme (Haase *et al.* 2010), (Nishimasu *et al.* 2012). Il existe certainement des étapes intermédiaires entre la transcription du précurseur et la prise en charge par Zuc mais elles restent inconnues à ce jour.

Ces piRNA sont ensuite dirigés vers des granules appelés les « Yb bodies », où se trouvent les protéines Armitage, Piwi, Shutdown, Vreteno et Yb. Il existe une autre structure, visualisée grâce à des expériences de RNA-FISH, connue sous le nom de « flam bodies » où les transcrits ainsi que les piRNA intermédiaires dérivant du locus *flamenco* s'accumulent. Les « flam bodies » forment un foci près des « Yb bodies » qui sont les sites de maturation des piRNA (Figure 23). Une mutation de l'activité RNA-binding de Yb détruit les Flam bodies ainsi que les Yb bodies. La protéine Yb pourrait aussi fixer les transcrits de *Flamenco* pour les diriger vers les flam bodies et les Yb bodies (Murota *et al.* 2014).

C'est dans les Yb bodies que les piRNA somatiques sont maturés et ensuite chargés sur Piwi grâce à la co-chaperone Shutdown. Les piRNA sont 2'-O-méthylés par la méthyltransférase Hen1 et subissent une étape de dégradation en 5' avant d'être exportés vers le noyau chargés sur Piwi pour réguler transcriptionnellement les ET cibles (Cf partie A.4.1 répression transcriptionnelle : TGS).

Les Yb bodies jouent un rôle fondamental dans la biogenèse des piRNA somatiques dont Yb est un des acteurs central. Lorsque la protéine Yb est mutée, on n'observe plus la formation des Yb bodies ni des flam bodies (Murota *et al.* 2014). La protéine Yb peut se fixer aux piRNA intermédiaires de flamenco et pourrait ainsi les diriger vers les flam bodies. En contexte mutant pour Yb, les piRNA ne sont plus chargés sur Piwi qui ne peut donc plus rentrer dans le noyau (Saito *et al.* 2010).

La littérature associée à la voie de biogenèse des piRNA est en rapide évolution. Deux nouvelles protéines ont été récemment identifiées grâce à des cribles génétiques, Minotaur (Vagin *et al.* 2013) et Gasz (Handler *et al.* 2013). Ces deux protéines se situent au niveau des mitochondries et interagissent avec Zuc. Leurs mutations impactent fortement la biogenèse des piRNA primaires.

Ce lien avec la mitochondrie semble important car Zuc est à la fois impliquée dans la biogenèse des piRNA somatiques et germinaux. D'autre part les nouveaux facteurs, Minotaur et Gasz, sont aussi en rapport avec cet organe. Cependant on ne connaît pas clairement à quel niveau l'activité mitochondriale entre en jeu lors de la biogenèse des piRNA qu'ils soient somatiques ou germinaux.

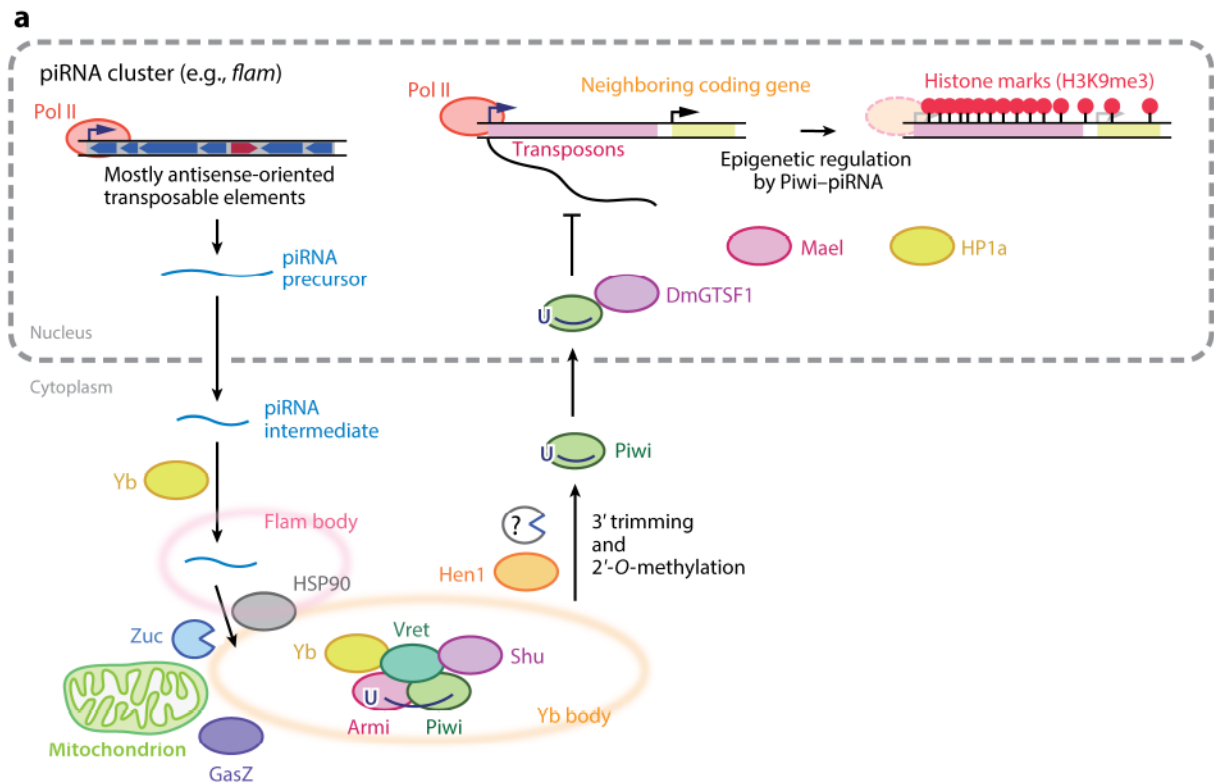


Figure 23. Biogenèse des piRNA somatiques et répression transcriptionnelle des ET. Les clusters producteurs de piRNA sont transcrits par la pol II et génèrent un précurseur qui est exporté dans le cytoplasme. Des molécules intermédiaires sont formées par des processus non connus. Zuc prend ensuite en charge les piRNA formés près des « Flam bodies » et « Yb bodies » où ces piRNA sont maturés et chargés sur Piwi. L'extrémité 3' est dégradée par une protéine inconnue et 2'-O-méthylée par Hen1. Le complexe Piwi-piRNA est exporté vers le noyau pour cibler les copies d'ET actives en déposant des marques chromatiques répressives comme H3K9me3. Un certain nombre de facteurs participent à ce processus comme DmGTSF1 (Astérix), Mael et HP1a (Iwasaki *et al.* 2015).

A.4 La répression des éléments transposables par les piRNA

A.4.1 répression transcriptionnelle : TGS

La protéine Piwi est à la fois cytoplasmique et nucléaire grâce au motif NLS présent dans sa partie N-terminale. De ce fait, seulement lorsqu'elle est chargée par son piRNA, elle est capable de rentrer dans le noyau et cibler les copies euchromatiques actives des ET (Figure 18). De récentes études ont pu montrer que Piwi peut réprimer l'expression des ET en bloquant leur transcription (Sienski *et al.* 2012), (Le Thomas *et al.* 2013). L'Histone 3 lysine 9 (H3K9me3) est méthylée sur les loci des ET induisant ainsi la formation d'hétérochromatine. Les piRNA permettent à Piwi de s'accrocher aux transcrits naissants d'ET et de recruter une histone méthyltransférase déposant la marque H3K9me3. Ceci entraîne la fixation de la protéine hétérochromatique HP1 empêchant ainsi la progression de l'ARN pol II. Cependant on ne sait pas encore par quels mécanismes la protéine Piwi chargée peut reconnaître précisément les copies d'ET à cibler (Figure 23).

Au moins deux protéines nucléaires ont récemment été identifiées comme co-facteur de Piwi, l'aidant dans sa fonction de répresseur transcriptionnel. La première est Astérix (aussi connue sous le nom de Gtsf1), c'est une protéine à doigt de zinc exprimée dans les ovaires (Donertas *et al.* 2013), (Ohtani *et al.* 2013). Il a été montré par immuno-précipitation qu'Astérix interagit physiquement avec Piwi. D'autre part, des analyses de ChIP-seq ont permis de mettre en évidence qu'en contexte mutant pour Astérix on observe une diminution de la marque H3K9me3 ainsi qu'une augmentation de la présence de l'ARN pol II au niveau des copies des ET. La deuxième a été identifiée de manière indépendante par deux laboratoires, Panoramix (Yu *et al.* 2015) ou aussi appelée Silencio (Sienski *et al.* 2015). La protéine Panoramix a été dirigée artificiellement contre le transcrit d'un ARN naissant d'un gène rapporteur ce qui a engendré d'une part la suppression de l'expression de ce gène et d'autre part le dépôt d'une marque hétérochromatique répressive H3K9me3 (Yu *et al.* 2015). Les mutants de ces deux protéines présentent les mêmes phénotypes qu'un mutant Piwi : dérégulation des ET, diminution de la marque H3K9me3 au site d'insertion des ET ainsi qu'au niveau des gènes situés à proximités des copies d'ET (Sienski *et al.* 2012). Ces deux protéines feraient donc partie de la machinerie de répression transcriptionnelle formée par Piwi dans le noyau.

Ces observations soulèvent une autre question qui a longtemps fait débat : quelle histone méthyltransférase dépose la marque H3K9me3 et comment cette marque est-elle maintenue au niveau des copies euchromatiques d'ET ? L'implication d'Eggless et d'HP1 a été suggérée

plusieurs fois (Rangan *et al.* 2011); (Huang *et al.* 2013) sans toutefois pouvoir attribuer un rôle précis à chacune des deux protéines dans la formation d'hétérochromatine. Sienski et ses collaborateurs semblent apporter un nouvel élément de réponse. La protéine Panoramix a été adressée au niveau du transcrit naissant d'un transgène rapporteur en contexte mutant pour Eggless et pour HP1. On observe une dérégulation du transgène rapporteur en contexte mutant pour Eggless et non pour HP1. D'autre part dans les mutants pour HP1, on peut observer une diminution d'un facteur 2 de la propagation de la marque H3K9me3. Les auteurs suggèrent alors qu'Eggless est requis pour la formation d'hétérochromatine et que HP1 participerait d'avantage au « spreading » de la marque H3K9me3 sur les copies d'ET ainsi que sur leur régions flanquantes (Sienski *et al.* 2015).

Toutes ces récentes découvertes ouvrent ainsi de nouvelles pistes quant à la compréhension de la répression transcriptionnelle des ET. Il faut toutefois noter que ces analyses que je viens de décrire sont effectuées en cellules OSS et OSC qui dérivent de la lignée ovarienne somatique où la voie des piRNA est plus simple que dans la lignée germinale.

A4.2 répression post-transcriptionnelle : PTGS

La répression post-transcriptionnelle ou PTGS (Post-Transcriptional Gene Silencing) a lieu dans le cytoplasme et plus précisément dans le nuage où l'on retrouve les protéines PIWI : Aub et AGO3. Ces protéines possèdent une activité d'endonucléase leur permettant de dégrader les transcrits d'ET complémentaires aux piRNA chargés par ces protéines. Lorsqu'un transcrit de copie active d'ET arrive dans le nuage, il est ciblé par un piRNA primaire complémentaire chargé par Aub et déclenche le mécanisme de ping-pong. Ceci entraîne sa dégradation et génère des piRNA complémentaires au transcrit de l'ET.

Ces deux mécanismes, TGS et PTGS, participent ensemble pour réprimer la transposition des ET dans la lignée germinale et donc agissent comme des gardiens de l'intégrité du génome. On peut toutefois se demander quelle est l'importance relative du TGS et du PTGS et aussi pourquoi la cellule a-t-elle mis en place deux processus distincts pour réguler la transposition des ET ? On peut imaginer que certaines copies d'ET puissent échapper au TGS dans le noyau et que le PTGS soit ainsi un second moyen de surveillance dans le cytoplasme.

A.5 L'activation d'un cluster pour la production de piRNA *de novo*

Le TSE a été mis en évidence par l'utilisation de transgènes silenciers télomériques insérés dans les TAS. Au cours de l'étude de ce processus, il a été mis en évidence qu'une lignée appelée *T-1* possédant 7 transgènes, *P-lacW* (transgènes possédant le gène *white*) non télomériques, insérés en tandem sur le bras droit du chromosome 2 est aussi un répresseur fort, alors que la lignée *BX2* qui possède les 7 mêmes transgènes insérés au même site cytologique n'est pas silencier (Ronsseray *et al.* 2001). La lignée *T-1* dérive de la lignée *BX2* suite à un traitement aux rayons X (Dorer and Henikoff 1994) ;(Dorer and Henikoff 1997). Ces deux lignées bien que possédant le même cluster n'ont pas les mêmes propriétés épigénétiques. Une hypothèse est qu'à l'issue du traitement aux rayons X, le cluster *T-1* est devenu capable de produire des piRNA ce qui n'est pas le cas de *BX2*. Une La question intéressante serait alors, peut-on activer un cluster inactif (*BX2*) par un cluster actif (*T-1*) par héritage maternel de piRNA homologues ?

Ainsi à mon arrivée au laboratoire, le phénomène de conversion épigénétique venait d'être mis en évidence (de Vanssay *et al.* 2012). En effet, lorsque le cytoplasme d'une femelle *T-1*, contenant des piRNA, est transmis maternellement à un locus *BX2* hérité paternellement, les descendants deviennent capables de produire des piRNA de *BX2* dès la 1^{ère} génération, G1. Cette conversion épigénétique est appelée paramutation. Ces individus ont été appelés *BX2** et sont maintenant capables de réprimer un transgène cible par TSE. Afin de confirmer ces observations, les petits ARN ovariens ont été testés par séquençage à haut débit. La souche *T-1* produit des piRNA alors que la souche *BX2* qui n'en produisait pas, devient capable à son tour d'en produire après la paramutation (*BX2**).

B. Régulation de gènes par les piRNA

La fonction la plus connue des piRNA est de réprimer la transposition des ET en lignée germinale cependant un nombre croissant d'articles semble décrire un rôle des piRNA dans la régulation de l'expression de gènes.

Je vous détaillerai dans cette partie deux cas de régulation spécifique de gènes par les piRNA chez la drosophile et je citerai deux exemples chez d'autres espèces.

B.1 Les 3'UTR, une source de piRNA dans les cellules somatiques

L'analyse de données de séquençage à haut débit de petits ARN dans les cellules OSC et OSS a permis de découvrir l'existence de piRNA dérivant des régions 3'UTR de certains ARN messagers (ARNm). Parmi les piRNA issus des ARNm, plus de 70% correspondent aux 3'UTR. Ce type de piRNA a pu être détecté aussi bien chez la drosophile que chez la souris et le xénope (Robine *et al.* 2009).

L'une des sources les plus importantes de ces « piRNA géniques » est le gène *traffic jam (tj)* (Saito *et al.* 2009),

(Robine *et al.* 2009). Ce gène code pour un facteur de transcription dans les cellules somatiques et est requis pour le développement des gonades mâles et femelles chez la drosophile. L'analyse des mutants pour *piwi* ainsi que pour *zuc* montre une forte diminution du nombre de piRNA produits. D'autre part, ces piRNA possèdent aussi un fort biais pour un U à la position 1 en 5'. Ces données suggèrent que ces piRNA seraient produits par la voie de biogenèse des piRNA somatiques. Ces observations soulèvent une question, quelle est la fonction de ces piRNA géniques et quelles sont leurs cibles ?

Les piRNA issus du 3'UTR *tj* sont issus du brin sens et ne peuvent donc cibler son transcrit. Ce gène ne semble donc pas être la cible de ces piRNA. Une fonction possible de ces piRNA pourrait être de réguler l'expression d'autres transcrits. Une recherche de gènes possédant une complémentarité de séquence avec les piRNA de *tj* a permis de mettre en évidence une cible potentielle, le gène *Fasciclin3 (Fas3)*. Ce gène code pour une molécule d'adhésion cellulaire. Des expériences de qRT-PCR ont montré que l'expression de *Fas3* est augmentée en contexte mutant pour *piwi* ou *tj* (Saito *et al.* 2009). Cependant on ne connaît pas le mécanisme exact d'action de ces piRNA sur leurs cibles potentielles.

D'autre part, le cas du gène *tj* est un peu particulier car celui-ci est un facteur de transcription dont l'une des cibles serait Piwi. En contexte mutant pour *tj* on perd l'expression de Piwi dans les cellules somatiques. Les auteurs proposent donc un modèle où *tj* permettrait d'une part d'activer l'expression de la protéine Piwi et d'autre part de produire des piRNA qui seraient pris en charge par Piwi pour réprimer ses cibles (Saito *et al.* 2009) (Figure 24).

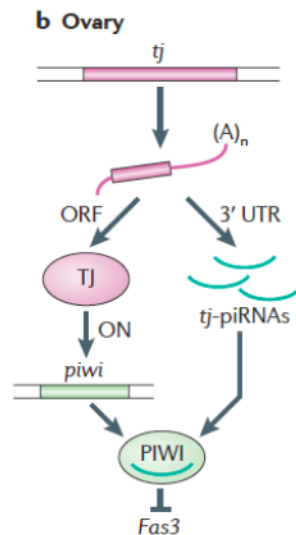


Figure 24. Régulation par *traffic jam* (*tj*) du gène *Fas3*. A partir du transcrit de *tj* on peut avoir la protéine Tj qui permet la régulation transcriptionnelle de *piwi* et d'autre part des piRNA issus du 3'UTR du transcrit qui peuvent être chargés par Piwi pour réguler des gènes dont *Fas3* (Siomi *et al.* 2011).

Qu'est ce qui définit certains 3'UTR de gènes en tant que régions productrices de piRNA ? Les travaux de Muerdter et collaborateurs en 2012 montrent que l'insertion artificielle d'une séquence codant pour la GFP dans le 3'UTR de *tj* permet de générer des piRNA correspondant à la GFP. Ces piRNA artificiels possèdent les mêmes propriétés que ceux du 3'UTR de *tj*. Ils sont orientés en sens et possèdent un biais de 1U en 5'. Cette observation suggère que la région entière du 3'UTR est reconnue pour la production de piRNA et surtout que cette région possède une information déterminante pour la spécifier comme cluster de piRNA. Très récemment, une étude a montré l'existence d'un élément de 100pb en *cis* du 3'UTR de *tj* qui serait suffisant pour induire la production de piRNA lorsqu'on y insère une séquence artificiellement. Les auteurs montrent aussi que la protéine Yb pourrait se fixer sur cet élément permettant de définir le transcrit issu de ce 3'UTR comme un précurseur pour la production de piRNA (Ishizu *et al.* 2015).

B.2 La régulation du gène *nanos* par les piRNA

nanos est un gène très étudié pour la biologie du développement embryonnaire chez la drosophile car il est déterminant pour spécifier le pôle postérieur (Lehmann and Nusslein-Volhard 1991). C'est un gène dont le transcrit est déposé maternellement dans le cytoplasme de l'ovocyte. Les ARNm de *nanos* présents dans l'embryon sont dégradés sauf au pôle

postérieur où ils sont traduits en protéines. Lorsque l'ARNm de *nanos* est traduit au pôle antérieur cela provoque des défauts de développement de la tête. Il existe donc ainsi un mécanisme permettant la régulation de l'ARNm de *nanos* dans le cytoplasme de l'embryon. La dégradation du transcrit passe par la machinerie de déadénylation impliquant les protéines CCR4 et Smaug (protéine de liaison à l'ARN). Il a été montré par des expériences de co-immunoprécipitations que les protéines Aub et AGO3 sont en complexe avec CCR4 et Smaug. On retrouve dans ce complexe des piRNA chargés par Aub et AGO3 issus des ET *roo* et *412* qui cibleraient la partie 3'UTR du transcrit par complémentarité de séquence (Rouget *et al.* 2010) (Figure 25). La voie des piRNA semble donc jouer un rôle dans la dégradation du transcrit de *nanos* dans la partie postérieure de l'embryon de drosophile. Ainsi en contexte mutant pour des gènes intervenants dans la biogenèse des piRNA, tels que Aub, AGO3 et Piwi on observe une augmentation du taux de transcrit de *nanos* qui n'est donc plus dégradé. Le rôle des piRNA de *roo* et *412* a été confirmé de deux manières. Premièrement, par des délétions des séquences complémentaires correspondants à ces deux ET dans le 3'UTR de *nanos* et deuxièmement, grâce à l'injection spécifique d'anti-piRNA pour *roo* et *412*. Ces deux expériences conduisent à une absence de déadénylation du transcrit de *nanos* ainsi qu'à des défauts de développement de la tête, un phénotype associé à la présence ectopique de *nanos* dans le pôle antérieur de l'embryon.

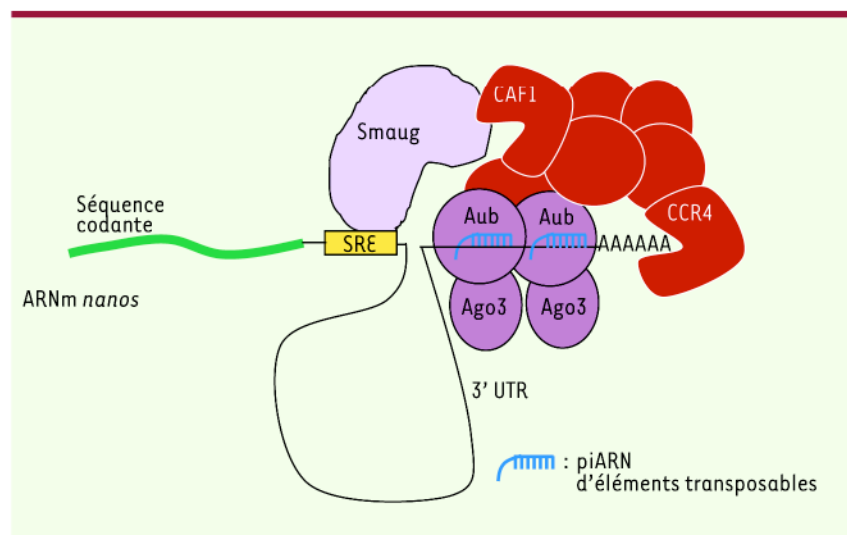


Figure 25. Modèle d'interaction ARN-protéine dans le complexe de déadénylation de l'ARNm *nanos*. Le complexe de déadénylation est formé par Smaug, CAF1 et CCR4. Les protéines Aub et AGO3 seraient en interaction avec ce complexe. Aub et AGO3 sont chargées par des piRNA d'ET complémentaires à la partie 3'UTR de l'ARNm de *nanos* (Papin and Simonelig 2011).

Plus récemment, un crible a été effectué par le même laboratoire grâce à la méthode d'iCLIP (Individual-nucleotide Crosslinking and Immunoprecipitation) afin de déterminer tous les transcrits maternels qui seraient dégradés par la même voie que *nanos* dans l'embryon. Cette technique permet de récupérer les ARN qui sont en interaction avec une protéine donnée, et est suivie d'une étape de séquençage à haut débit de ces ARN. Cette étude a ainsi permis de mettre en évidence au moins 182 gènes dont les transcrits sont déposés maternellement et dont la dégradation serait associée à Aub chargée par un piRNA (Barckmann *et al.* 2015). Ces résultats suggèrent donc un rôle important pour la protéine Aub et les piRNA dans la régulation de l'expression de gènes dans l'embryon précoce.

B3. Régulation de gènes par les piRNA chez d'autres espèces

B.1.1 La souris

Chez la souris, il existe trois protéines PIWI : MIWI, MILI et MIWI2 qui sont respectivement les homologues des protéines Piwi, Aub et AGO3 chez la drosophile. Les piRNA sont produits en deux temps durant la spermatogenèse, les piRNA pré-pachytènes dérivant d'ET et les piRNA pachytènes provenant de plusieurs clusters qui sont des régions intergéniques. La fonction des piRNA pachytènes reste inconnue car on ne retrouve aucune séquence qui leur est complémentaire à part les régions dont ils sont issus (Pour revue : (Iwasaki *et al.* 2015)). Un rôle des piRNA pachytènes a été proposé dans la régulation de gènes. Ces piRNA chargés par MIWI permettraient la dégradation de certains ARNm dans les spermatides (Gou *et al.* 2014). Ces piRNA possèdent des séquences partiellement complémentaires dans les régions 3'UTR de leur ARNm cibles. De plus, MIWI interagit avec la protéine CAF1, membre du complexe de déadénylation CCR4-NOT, et l'expression de plus 5000 gènes augmente après un *knockdown* de MIWI ou CAF1. Parmi ces gènes, 90% sont communs aux deux contextes mutants. Les transcrits cibles sont donc reconnus par la machinerie de déadénylation où se trouve MIWI chargée par son piRNA entraînant leur dégradation. La faible complémentarité des piRNA pachytènes permettrait donc de cibler un large répertoire de gènes.

B.1.2 Le ver à soie

Une étude récente a permis de mettre en évidence le rôle des piRNA dans la détermination du sexe chez le ver à soie, *Bombyx mori* (Kiuchi *et al.* 2014). Un déterminant connu jusqu'à présent est le gène *Bmdsx* (*Bombyx mori Doublesex*). L'épissage alternatif de ce gène spécifie l'expression de facteurs mâle ou femelle. Cependant le processus contrôlant cet épissage n'a pas été clairement identifié. Chez la plupart des lépidoptères, le sexe dépend des chromosomes WZ. Les embryons possédant un chromosome W et Z seront des femelles alors que ceux possédant deux chromosomes Z seront des mâles. Le chromosome W est majoritairement composé d'ET permettant la production de piRNA. Un gène répété en tandem *Fem* (*Feminizer*) qui est un précurseur de piRNA a été retrouvé sur le chromosome W. Les auteurs montrent notamment que la cible de ces piRNA est un gène appelé *Masc* (*Masculinizer*), codant pour un facteur nécessaire pour l'épissage de *Bmdsx*. Chez les mâles, le gène *Masc* est exprimé et conduit à l'épissage de *Bmdsx* spécifique des mâles alors que chez la femelle *Masc* est réprimé par les piRNA issus du gène *Fem*. Dans ce cas, c'est la forme épissée de *Bmdsx* spécifiant le sexe féminin qui est produite (Figure 26).

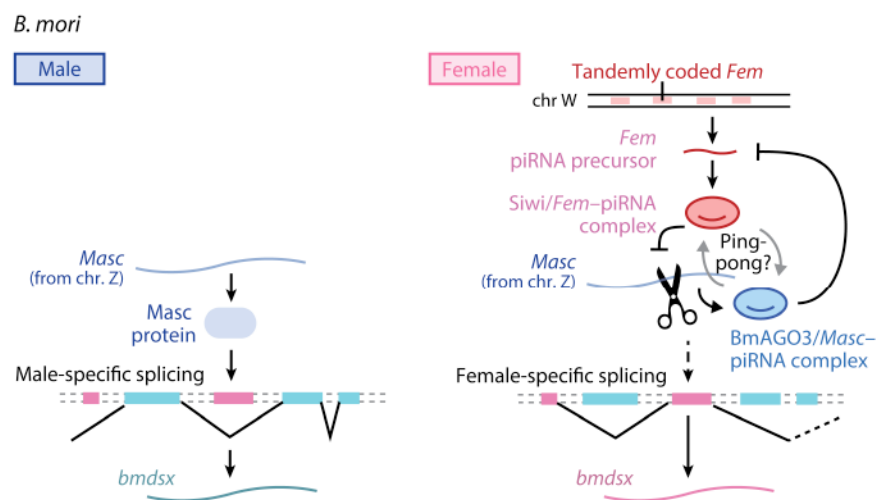


Figure 26. Détermination du sexe par les piRNA chez *Bombyx mori*. Les piRNA issus du gène répété en tandem *Fem* peuvent cibler le transcrit issu du gène *Masc*. Ce dernier étant dégradé, c'est la version épissée de *Bmdsx* induisant l'expression de gène spécifique de la féminisation qui est exprimée. Dans le cas contraire, si *Masc* est exprimé, c'est la version épissée de *Bmdsx* induisant l'expression de gène spécifique de la masculinisation qui est exprimée (Iwasaki *et al.* 2015).

B.1.3 L'aplysie

Enfin je vous cite un dernier exemple chez l'aplysie où les piRNA réguleraient un gène, *CREB2* qui code pour un répresseur transcriptionnel inhibant l'expression de gènes impliqués dans la mémoire à long terme (Rajasethupathy *et al.* 2012). Dans cette étude, les auteurs ont découvert la présence de piRNA en dehors de la lignée germinale, dans le système nerveux central de l'aplysie. Ces piRNA co-immunoprécipitent avec la protéine Piwi. Un seul piRNA, piR-F est impliqué dans ce processus de répression. Le complexe Piwi-piR-F induirait la méthylation du promoteur de *CREB2* ce qui permettrait la répression transcriptionnelle de ce gène (Figure 27).

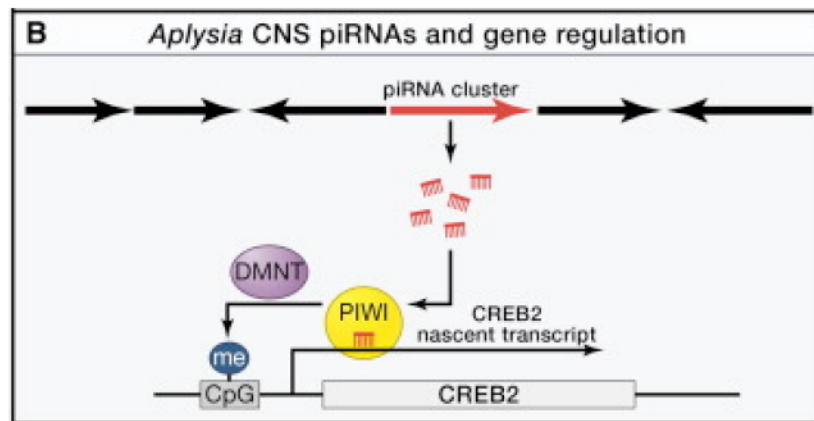


Figure 27. Régulation du gène *CREB2* par les piRNA. Les complexes Piwi-piRNA peuvent induire la méthylation (me) des îlots CpG par l'ADN méthyltransférase (DMNT pour DNA methyltransferase) (Moazed 2012).

Ces exemples nous montrent qu'en dehors de leur rôle de répression d'ET, les piRNA peuvent aussi réguler l'expression de gènes codants par des mécanismes variés. Cette régulation peut se faire aussi bien à un niveau transcriptionnel que post-transcriptionnel.

C. Objectifs de l'étude

En résumé, l'étude de l'élément *P* chez la drosophile a permis de mettre en évidence un point chaud d'insertion de cet élément qui est le site cytologique 1A et plus particulièrement le locus *TAS-X*. Une fois l'élément *P* inséré dans ce locus cet élément acquiert la capacité de réprimer la transposition des éléments *P* présent ailleurs dans le génome. Il a été montré par la suite que cette répression passe par une classe particulière de petits ARN, les piRNA. Il existe dans le génome de *D. melanogaster* plus de 140 locus producteurs de piRNA (dont le locus TAS) qui répriment de façon plus générale l'expression des éléments transposables dans les gonades.

Au cours de cette thèse, je me suis intéressée plus particulièrement à l'étude d'un locus producteur de piRNA, le locus TAS. En effet, celui-ci possède un statut particulier par rapport aux autres locus. Il est situé dans les régions sub-téломériques et est polymorphique car il peut être perdu lors d'événements aléatoires de cassures téломériques au cours des générations. J'ai donc consacré ma thèse à la réalisation de deux objectifs. Premièrement, j'ai étudié de manière plus fine la structure de ce locus ainsi que son polymorphisme au sein des populations naturelles par rapport aux populations de drosophiles présentes dans les laboratoires. La deuxième partie du projet concerne l'analyse du rôle fonctionnel du locus TAS. En effet, ce locus fait parti des locus producteurs de piRNA. J'ai étudié plus particulièrement le rôle de ce locus lorsqu'un élément transposable s'y insère, c'est à dire comment se met en place les mécanismes permettant sa répression au cours des premières générations. Enfin, j'ai pu initier une étude portant sur l'hypothèse du rôle des piRNA produits par locus TAS dans la répression de l'expression de gènes en lignée germinale.

Cette étude axée sur un locus original producteur de piRNA a permis de mettre en évidence d'une part que la structure particulière de ce locus peut refléter son histoire évolutive au sein du groupe *melanogaster*. D'autre part ce locus possède des propriétés épigénétiques particulières qu'il peut transmettre aux séquences qui s'y insèrent permettant l'établissement de leur répression au sein du génome.

II-Résultats

1. Article I

Présentation l'article I

Le premier objectif de cette thèse était de caractériser structurellement le locus TAS et d'étudier sa distribution au sein des souches de drosophiles de laboratoires comparée aux drosophiles présentes dans la nature. En effet, il y a longtemps Bob Levis (communication personnelle) avait suggéré que dans les laboratoires, les séquences TAS pouvaient être perdues. Par ailleurs, au laboratoire, il a parfois été observé une perte de la répression de transgènes cibles *P-lacZ* par les transgènes *P-lacZ* télomériques ainsi qu'une perte de la répression de l'élément *P* dans certaines souches. Cette observation était accompagné de l'hypothèse que cette perte de répression pouvait être due à la perte des séquences TAS où sont justement insérés les transgènes *P-lacZ* télomérique ainsi que ces élément *P*. Afin de vérifier cette hypothèse, des souches de drosophiles ont été collectées dans la nature entre 2013 et 2015 en Europe, en Amérique du Sud et en Afrique. Ces souches ont ensuite été analysées par des expériences d'hybridation *in situ* combinées à des PCR sur plusieurs générations afin de suivre le statut du *TAS-X*. Par ailleurs, la structure des locus TAS ainsi que la production de piRNA a été étudiée grâce à des analyses bio-informatiques. En effet, la famille des TAS-R est constituée de plusieurs domaines répétés non codants et se distingue de la famille des TAS-L par la présence de trois soloLTR de l'élément *Invader 4* retrouvés dans chaque unité de répétition. La structure des domaines ne correspondant pas à *Invader 4* a été étudiée plus particulièrement. Une hypothèse est que ces domaines pourraient avoir une origine commune.

Cette étude sera très prochainement soumise pour une publication dans la revue DNA RESEARCH

Article I

Telomeric Associated Sequences in Drosophila: structure, dynamic and properties of a piRNA cluster

Telomeric Associated Sequences in *Drosophila*: structure, dynamic and properties of a piRNA cluster

Amna Asif-Laidin¹, Valérie Delmarre¹, Jeanne Laurentie¹, Wolfgang J. Miller², Stéphane Ronsseray¹ and Laure Teyssset¹

¹Sorbonne Universités, UPMC University of Paris 06, Institut de Biologie Paris-Seine, UMR7622, Laboratoire Biologie du Développement, F-75005, Paris, France CNRS, UMR7622, "Epigenetic Repression and Mobile DNA," F-75005, Paris, France, ²Lab Genome Dynamics, Dept Cell & Developmental Biology, Center of Anatomy and Cell Biology, Medical University of Vienna, A-1090 Vienna, Austria.

Email: laure.teyssset@upmc.fr

Running title:

Subtelomeric piRNA cluster characterization

20 **Abstract and key words**

21

22 Telomeric Associated Sequences (TAS) are sub-telomeric non-coding heterochromatic repeats. TAS-R found
23 on *X*, *2R* and *3R* and TAS-L found on *2L* and *3L* chromosomal arms are the two major families of TAS
24 identified in *D. melanogaster*. In this study we revisited their structure and dynamics. By performing *in silico*
25 analysis, the TAS-R family obviously consists of six domains placed in a specific order on a given telomere.
26 One of them, the domain F specific of *TAS-X*, contains a region called *TLL* that appears to be the ancestor of
27 TAS-R and TAS-L. TAS clusters have been shown to play a major role in the establishment of *P* element
28 repression following invasion of *D. melanogaster* genome. It involves piRNA production from TAS-R loci. We
29 also show the piRNA profiles produced by TAS-R and TAS-L loci, plus their orthologs identified in
30 *D. sechellia* and *D. simulans*. An extended population analysis revealed that *D. melanogaster* strains always
31 possess at least 2 TAS-R loci. Surprisingly, *TAS-X* is highly maintained in flies living in the wild and rarely
32 retain in lab flies, suggesting that a yet unknown selection pressure exists in natural environment that is
33 alleviated after about 2 years of domestication in laboratory conditions.

34

35 Key words: TAS, *Drosophila*, piRNA, Transposable Element, domestication

36

Introduction

In *Drosophila melanogaster* genome, Telomeric Associated Sequences (TAS) are moderately repeated sequences that correspond to non-coding heterochromatin. They lie between euchromatin and the proximal telomeric transposon arrays, consisting on specialized non-LTR retrotransposons: *HeT-A*, *TAHRE* and *TART* (HTT) (Karpen and Spradling 1992, Levis, Ganesan et al. 1993, Walter, Jang et al. 1995, Biessmann and Mason 2003, Abad, De Pablos et al. 2004, Mason and Villasante 2014). Due to their position, they constitute a transition zone between telomeres and genes. The initial molecular analysis of TAS locus from the *X* chromosome (*TAS-X*) has been derived from the study of a 10kb fragment isolated from the *Dp1187* mini-chromosome (Karpen and Spradling 1992). The telomeric region of this modified chromosome contains a *HeT* member followed by a tandem of four TAS repeats of 1.8 kb for the longest and 2 repeats of 0.9kb. The four TAS repeats are non-coding units (Karpen and Spradling 1992) (**Fig. 1**). Each repeat contains three tandem subrepeats of 170 bp long that has been found later to be derived from a solo-LTR of the *Invader4* (*INV-4*) retrotransposon (Quesneville, Bergman et al. 2005).

Early *in situ* hybridization studies on polytene chromosomes and recent sequence alignments identified common regions restricted to *INV-4* between *TAS-X*, *TAS-2R* and *TAS-3R* (Levis 1989, Karpen and Spradling 1992, Yin and Lin 2007). These three subtelomeric regions have been classified in a group called TAS-R family to easily differentiate them from the TAS-L family localized on *2L* and *3L* telomeres (Silva-Sousa and Casacuberta 2013). TAS-L are made of shorter non coding repeats of 461bp (Walter, Jang et al. 1995). The TAS-R and TAS-L family were considered to share no similarity (Levis, Ganesan et al. 1993, Biessmann and Mason 2003) up to a recent review mentioning common motifs (Mason and Villasante 2014).

TAS-R family has been shown to play a major role in the establishment of *P* element repression following invasion of *D. melanogaster* genome between the 1950's and the 1970's (Anxolabehere, Kidwell et al. 1988, Ronsseray, Lehmann et al. 1989, Ronsseray, Lehmann et al. 1991). The study of this repression led to the discovery of an ovarian silencing mechanism involving piRNA biogenesis called *Trans*-Silencing Effect (TSE) (Roche and Rio 1998, Ronsseray, Josse et al. 2003, Josse, Teyssset et al. 2007). piRNA are a class of small RNA of 23 to 29 nt produced in germline from piRNA producer loci (Brennecke, Aravin et al. 2007). These loci are heterochromatic, mainly made of fragments of Transposable Elements (TE), and involved in repression of transposition of TE through transcriptional and/or posttranscriptional level. TAS-R, especially *TAS-X*, appears to be piRNA producer loci, where *P* element (as well as *P*-derived transgenes) can insert. Once *P* is inserted, TAS locus will also produce piRNA from *P* (Brennecke, Malone et al. 2008, Todeschini, Teyssset et al. 2010, de Vanssay, Bouge et al. 2012, Muerdter, Olovnikov et al. 2012). Doing so, euchromatic *P* copies will be repressed in germline leading to the so-called *TSE*. It is noteworthy that the majority of *P* element or *P*-transgene insertions possessing repression capacities have been identified in *TAS-X*, only few were found inserted in *TAS-2R* and *TAS-3R*, and those latter were only *P*-transgenes.

Current literature studying repression of cognate TE is dealing with the fact that each element can be regulated by multiple piRNA clusters in the same time. Here we present a study of a specific piRNA cluster,

i.e., the *TAS-X* located on the tip of the *X* chromosome, which as we found appears to possess a unique domain (domain F). We further present data emphasizing that the three TAS-R units are more closely related to each other than previously thought. Each TAS-R unit is made of domains that are scattered in a specific order for each subtelomeric chromosome. By a population analysis, we analysed the dynamics of the TAS-R family in term of presence on each telomeres as well as the structure of the repeats. We focused on the *TAS-X* locus by studying its distribution within strains fixed in laboratory for decades or recently collected from the wild. As it was previously suspected (Levis RW, unpublished data), we showed by a population analysis that *Drosophila melanogaster TAS-X* are under some selection pressure in the wild that maintain them, and that selection pressure is abolished when strains are kept for a few years under laboratory conditions. Finally thanks to sequence similarities found between domain in *TAS-X* and TAS-L that we called *TLL*, we were able to identify the orthologue of *D. melanogaster TAS* in sister species, *i.e.* *D. sechellia*, *D. simulans* and *D. erecta*. Combining *in silico*, *in situ* hybridization and small RNA sequencing, we pinpointed that these sequences are telomeric, repeated and genuine piRNA producer loci that can be involved in TE repression.

Taken together, TAS appear to be ancestral subtelomeric sequences that evolved to give rise to the two families in *D. melanogaster* (TAS-R and TAS-L). These loci and more specifically *TAS-X* played a fundamental role during *P* invasion at least in the *D. melanogaster* genome. The surprise of our results comes from the fact that the *TAS-X* locus seems absolutely required in wild life but can be lost under laboratory conditions indicating a yet unknown function essential in wild environment. TAS locus is of particular interest because it concerns a unique situation allowing to study the biology of piRNA production coming from a single locus against a peculiar TE, the *P* element that invaded recently *Drosophila* genome, providing strains with lots of copies or totally devoid of copy.

Materials and Methods

Fly stocks of *Drosophila* and related species Strains

For this study, we used the following Fly stocks: *y^l; cn^l bw^l sp^l* (FBst0002057), *Canton-S* (FBst0000001), *Oregon-R-C* (FBst0000005), *VAG_2* (FBst0003876), *Wild_11C* (FBst0003894), *BER_2* (FBst0003840) and *Wild_5A* (FBst0003885) are from Bloomington stock center. The *w¹¹¹⁸*, *P-1152*, *Na-P(1A)* and *Lk-P(1A)* are our laboratory stocks. *P-1152* carries two *P-lacZ* fusion enhancer trap transgene (FBti0005700) inserted in *TAS-X* that contains an in-frame translational fusion of the *E. coli lacZ* gene to the second exon of the *P* transposase gene, as well as the *rosy+* gene as a transformation marker (O'Kane and Gehring 1987). The *Lk-P(1A)* line carries two full-length *P* elements inserted in *TAS-X*. *Na-P(1A)* carries a *P* element inserted in *TAS-X* that has undergone a telomeric deletion leaving a 5' deleted *P* element. These telomeric insertion have been genetically isolated from inbred lines derived from wild populations collected in Lerik, Azerbaijan for *Lk-P(1A)* (Ronsseray, Lehmann et al. 1991) and from Nasrallah, Tunisia for *Na-P(1A)* (Marin, Lehmann et al.

2000). *Drosophila sechellia* strain was kindly provided by Aurélie Hua-Van (EGCE, CNRS-IRD-Univ.Paris-Dud- Univ Paris Saclay).

We recently collected *Drosophila* population from wild in Europe (France and Austria), Africa (Ethiopia and Senegal) and South America (Brazil and French Guiana) between 2013 and 2015. The strains collected in Marsais in 2013 (France) and in Gotheron in 2012 (Valence, France) were kindly provided by Jean-Michel Gibert and Cristina Vieira respectively (Supplementary Table 3).

Sequence accession numbers

The following sequences were used in this study: *TAS-X* (L03284) (Karpen and Spradling 1992), *TAS 2L* (U35404) (Walter, Jang et al. 1995), *Canton-S* cosmid (AL031884) (Benos, Gatt et al. 2001), clone AF520587 (Boussy and Itoh 2004), *Invader 4* (FBgn0063427) and *TAS-2R* and *TAS-3R* were recovered from Lin and Yin, 2007 (Yin and Lin 2007), *y^l*; *cn^l bw^l sp^l* from the FlyBase (Adams, Celniker et al. 2000).

Preparation of polytene chromosomes and fluorescence in situ hybridization

Salivary glands from third instar larvae grown at 18°C were dissected and fixed in 45% acetic acid. After squashing, slides were quick-frozen in liquid nitrogen and dehydrated in ethanol (90%). Slides were then washed 30 min in 2X SSC at 65°C and dehydrated in ethanol baths (90%). Before hybridization, chromosomal DNA was denatured in NaOH 0.07N. Biotinylated DNA probes were made using BioNick Labeling System (Invitrogen) and Digoxigenin probes with DIG-Nick Translation Mix (Roche). Slides were incubated with 1µg of probe at 37°C overnight. Slides were washed two times in 2X SSC at 37°C. For detection, we used FITC-extravidine at 50µg/mL (Sigma) and anti-DIG-rhodamine at 4µg/ mL (Roche) final concentrations. Slides were stained in DAPI-VectaShield mounting medium and observed using Leica TCS SP5 microscope system.

Sequence and phylogeny analysis

Sequence comparisons were performed using BLAST (NCBI) and SeaView using MUSCLE for multiple sequence alignments. Minor manual adjustments of gaps were done if needed (Gouy, Guindon et al. 2010).

DNA extraction and PCR amplification

One to fifty flies were manually homogenized in a buffer containing 0.1M Tris pH9, 0.1M EDTA pH8 and 1% SDS and then incubated at 70°C for 30 min. Potassium Acetate 8M was added prior centrifugation, and DNA was precipitated with isopropanol and wash in Ethanol 70%. Pellet of DNA was resuspended in water and PCR was performed using primers listed in **Supplementary Table 1**. Since no PCR amplification could be due to poor DNA quality, we confirmed the amplificability of DNA using other primers (data not shown).

Small RNA sequencing

Total ARN extraction, small RNA libraries preparation and sequencing were performed as described in Hermant et al (Hermant, Boivin et al. 2015). Sequence reads in FASTQ format were trimmed from the adapter sequence 5'-TGGAATTCTCGGGTGCCAAG-3' and matched to reference sequences (TAS, *P-element* or *D. sechellia* repeats) using Bowtie (Langmead, Trapnell et al. 2009). Only 23-29nt reads matching the reference sequences with 0 mismatch were retained for subsequent analysis. Annotation of small RNA libraries is described in **Supplementary Table 2**. Sequence length distributions and small RNA mapping were generated from bowtie alignments using Python and R (www.r-project.org/) scripts, which were wrapped and run in a Galaxy instance publicly available at <http://plastissipi.fr>. Tools and workflows used in this study can be downloaded from this Galaxy instance.

3. Results and Discussion

3.1. Structural analysis of TAS-R vs TAS-L family arrays

X, 2R and 3R chromosomal arms carry subtelomeric TAS-R sequences called *TAS-X*, *TAS-2R* and *TAS-3R*, respectively. They are characterized by a series of subrepeats corresponding to *INV-4* sequence. Since they appear to be the main loci in the *Drosophila* genome involved in the repression mechanism of *P* element transposition, we aimed to revisit the structure and the sequence composition of the TAS-R loci in the *Drosophila melanogaster* genome in the light of the sequence data available today. The *Dp1187* minichromosome was originally used to describe the structure of subtelomeric region of the *X* chromosome. It contains a *HeT* member, a telomeric Transposable Element (TE), followed by a tandem of four TAS repeats (a *TAS-A* repeat of 205 bp, 2 units of 1872 bp and 1858 bp called *TAS-B* and *C* respectively, and a *TAS-D* repeat of 1575 bp, sharing 95% of similarity with *TAS-B* and *D*) and two 0.9kb-repeats (Karpen and Spradling 1992).

We first performed multiple alignment comparisons using the four TAS repeated units (**Fig. 1A and Supplementary Fig. 1**). As previously described, the *TAS-A* is a deleted repeat most likely generated by a telomeric 5' terminal deletion healed by *Het-A* transposition event (92% identity with *Het-A* canonical sequence, data not shown). As presented above, while *TAS-B* only differs from *TAS-C* by few indels, *TAS-D* presents a deletion of 313 bp within the 3' region of the unit (**Supplementary Fig. 1**). Given the fact that the 1.8kb *TAS-B* unit is the longest, it was considered to be the reference for *TAS-X* repeat.

Next we revisited the internal subrepeats known to be a 172 bp sequence of solo LTR of *INV-4* that are hot spot for *P* element insertion (Karpen and Spradling 1992, Josse, Maurel-Zaffran et al. 2008). *INV-4* is a LTR-containing retrotransposon present in nine euchromatic copies of which 2 are full length in the annotated *D. melanogaster* genome (Kaminker, Bergman et al. 2002). We aligned the LTR of the canonical *INV-4* with *TAS-B* sequence (**Supplementary Fig. 2 A and B**). This analysis allowed us to pinpoint that the three

subrepeats of 172 bp correspond to central part of the LTR. This sequence comparison allowed us to uncover that the portion corresponding to the LTR is longer than originally thought. The first subrepeat covers the LTR from the 172 bp core up to the 3' end (missing the last 16 nucleotides), while the third subrepeat covers from the first nucleotide to the core. This results in subrepeats of 248, 172 and 235 bp that are 84, 83 and 88% identical to the canonical LTR respectively. It was known that TAS sequences could be detected at the tip of the 2R and 3R chromosome arms mostly because of the presence of *INV-4* subrepeats (Levis 1989, Karpen and Spradling 1992, Yin and Lin 2007). The canonical *TAS-2R* and *TAS-3R* units described are 1002 bp and 982 bp long respectively (**Fig. 2**). Using the same approach of sequence comparison, the three subrepeats identified in the *TAS-3R* exhibit a similar organization to that one found in *TAS-X* with a shorter middle subrepeat. Two middle shorter subrepeats are also found on *TAS-2R* uncovering a fourth subrepeat of the *INV-4* LTR that was not previously described (**Supplementary Fig. 2 and Fig. 2**). Therefore, the portion of *INV-4* LTR within the TAS loci is longer than previously thought. LTR sequences are present on retroelements sequences, including Retroviruses and LTR-containing Retrotransposons. They are involved in reverse transcription, contain enhancers, promoters, and terminal of transcription signals. Therefore, the finding that all of these regulatory sequences are present in the TAS locus could be of interest when considering TAS as a piRNA cluster that need to be transcribed.

We then asked whether the shared origin between the *TAS-X*, *TAS-2R* and *TAS-3R* was restricted to the *INV-4* sequences. By performing multiple alignments, we identified five other domains arranged in a specific order for each chromosome (**Fig. 2**). These domains were named domain C, D, E, F and G. While domains E and D are present in all three chromosomal TAS loci, domain C is specific for the *TAS-X* and *TAS-3R* and domain G specific for the *TAS-2R* and *TAS-3R*. Only one domain of 909 bp called domain F, was found to be restricted to the *TAS-X* subtelomeric region. Through the identification of domains, we unveiled that the three TAS loci are more closely related to each other than originally thought. This allowed us to emphasize a common ancestor sequence made of at least six modules, including *INV-4* LTR assembled in a specific order that was fixed specifically on each telomere.

We then investigated the possible origin of each of the domains identified above. Searches did not reveal any similarities with known sequences for the domains E, C, D and G either in the *Drosophila* reference genome from the *y¹; cn¹ bw¹ sp¹* strain or in other sequences available in DataBases. BLAST searches using the domain F identified a 158 bp region that exhibits similarities with two domains of the proximal 0.9kb repeats that followed the four TAS repeats on the mini-chromosome *DP1187* (**Fig. 1A and Fig. 2D**). BLAST searches using the *D. melanogaster* reference genome identified that the same 158 bp domain is also present in the TAS-L family, i.e. *TAS-2L* and *TAS-3L* (**Fig. 2E**). *TAS-2L* and *TAS-3L* are identical telomeric minisatellites repeats of 461bp (99% identity, data not shown) (Walter, Jang et al. 1995). Eleven and 43 full length repeats are found on the subtelomeric of 2L and 3L respectively in the *y¹; cn¹ bw¹ sp¹* reference genome (data not shown). It was previously assumed that the TAS-R and TAS-L families do not share common sequences, up to a recent review

mentioning identities between the two families (Mason and Villasante 2014). Here, we unveiled that the TAS-L minisatellite is made of about six tandemly repeated, but incomplete sequences of 121, 54, 29, 48, 45 and 20 bp with identity of about 80% with the 158 bp domain identified in domain F (Fig. 2E). For this reason, we named this 158 bp domain, *TAS-L* like (*TLL*) domain (Fig. 2). It has to be mentioned, however, that *TLL* is absent from the *TAS-D* repeat since it is included in the 313 deleted base pairs described above (Supplementary Fig. 1). BLAST searches using the whole domain F also identified similarities between *TLL* and an unmapped clone isolated from a strain collected in Samarkand, Uzbekistan (73% of identity), composed of fragments of TE including *ninja*, *Tirant/ZAM*, and *Hobo* (Boussy and Itoh 2004) (data not shown). Such organization was not found in the reference genome and could be part of an unknown heterochromatic piRNA cluster (data not shown). Identification of *TLL* among a genomic region bearing several fragments of TE could also predict a TE origin for *TLL*. The *TLL* domain was used to search in the annotated genome and two similar domains of about 80% identity were found in centromeric region.

To sum up this analysis revealed that the main Telomeric Associated Sequences in *D. melanogaster* have a common ancestor composed of the 158bp-*TLL* domain that evolved and lead to the emergence of the TAS-R family, located on *X*, *2R*, and *3R*, and of the TAS-L family, located on *2L* and *3L*. Hence, these two families of telomeric repeats are more closely related than previously anticipated. The fact that subtelomeres are made of domains in a specific order depending on the chromosome seems to indicate that for some reasons no gene conversion can occur between the TAS-R family loci. This apparent repression of gene conversion is not observed on the TAS-L family suggesting different dynamic between the two TAS families.

3.2. Distribution of TASR family in Laboratory vs Wild population

Previous *in situ* hybridization studies have shown that the 1.8-kb *TAS-X* probe predominantly marks telomeres (Karpen and Spradling 1992). The genome sequence of *y^l; cn^l bw^l sp^l* revealed that TAS sequences are absent from the *X* telomere (cytological 1A site), whereas we identified one complete repeat on *2R* and six complete repeats on *3R* telomeres (data not shown). Since the sequence of chromosomes *2R* and *3R* does not include the telomeric arrays, it is possible that the number of TAS repeats on autosomes is higher. Molecular analysis identified *P* elements or *P*-transgenes inserted in *TAS-X* in number of strains, for instance *Lk-P(1A)* strain possess two complete *P* elements (Ronsseray, Lehmann et al. 1991), *NA-P(1A)* contains a 5' deleted *P* element (Marin, Lehmann et al. 2000) and *P-1152* strain possess two *P*-transgenes in *TAS-X* (Josse, Maurel-Zaffran et al. 2008). These strains are maintained in our laboratory for these insertions, otherwise they will probably lose the insertions, as it was the case for the *P-1152* strain maintained by the Bloomington Stock Center that kept *TAS-X* units but lose the *P*-transgenes inserted in the *TAS-X* (data not shown). It appears that the subtelomeric arrays can be easily also lost without a strongly deleterious effect (Levis 1989, Mason and Villasante 2014). We hence decided to focus on TAS-R family, especially *TAS-X*, by conducting an extended analysis on their

distribution among different strains to ask whether these differences exist, if they are random or triggered by some kind of selection pressure.

We first started to test the distribution of TAS-R on the 5 unrelated standard laboratory strains studied for almost a century for the oldest, *i.e.* *Canton-S*, *Oregon-R-C*, *w¹¹¹⁸*, *yw^c* and *y^l; cn^l bw^l sp^l*. To address this question, we analyzed by *in situ* hybridization on polytene chromosomes of third instar larvae using a 1.8-kb TAS probe. Control experiences using *y^l; cn^l bw^l sp^l* show staining on the *TAS-2R* and *-3R*, but none on *TAS-X* as expected for the reference genome in FlyBase (**Table 1** and data not shown). We also tested this probe on *Lk-P(1A)* and *P-1152* strains that possess *TAS-X*. It shows strong labels on the *X* (1A) and *2R* (60F) for *Lk-P(1A)* strain (**Fig. 3** and **Table 1**), and *X* (1A) and *3R* (100F) for *P-1152* strain (**Supplementary Fig. 3** and **Table 1**). Therefore due to the number of shared domains between members of the TAS-R family, this probe appears capable to label TAS repeats on *X* as well as on *2R* and *3R* telomeres, but not on *TAS-2L* and *TAS-3L*. We also verified that the experimental conditions were sensitive enough to be able to detect an euchromatic transgene containing a 1.2 kb TAS unit (data not shown, (Boivin, Gally et al. 2003)). **Fig. 3, Supplementary Fig. 3** and **Table 1** show that all the reference strains (*y^l; cn^l bw^l sp^l*, *Canton-S*, *yw^c*; *w¹¹¹⁸* and *Oregon-R-C*) contain TAS on autosomal loci *i.e.* *2R* and *3R* and none on the *X* chromosome, except for *Oregon-R-C*. It is noteworthy that Oregon strain shows polymorphism depending on stock considered; ours possess the 3 TAS loci, while other stocks have been shown to be devoid of the *TAS-X* locus (Andreyeva, Belyaeva et al. 2005). We confirmed the results of presence or absence of *TAS-X* by performing PCR experiments using specific primers for the domain F included in the *TAS-B* unit (**Table 1** and **Supplementary Fig. 1**). These PCR results showed that PCR amplifications are consistent with the *in situ* hybridization results (**Table 1**), given us an easy assay to test the presence of *TAS-X* in a strain. We also performed this experiment on *NA-P(1A)* suffering from a telomeric breakage in 5' region of a *P* element inserted in a *INV-4* subrepeat that was healed by a transposition event of *Het-A* (Marin, Lehmann et al. 2000). The *in situ* labelling shows a faint band on *X* telomere (**Supplementary Fig. 3** and **Table 1**). This result indicates that our experimental conditions are sensitive enough to highlight one deleted unit. Despite the *in situ* labelling, no PCR amplification was obtained. This apparent discrepancy can be easily explained because the *P*-element is inserted in the last TAS unit, *TAS-D*, that is the most divergent from the canonical repeat used to design primers (**Supplementary Fig. 1**). PCRs, with specific primers for *TAS-D* unit (**Supplementary Fig. 1**), confirmed that at least one *TAS-D* repeat is present (**Supplementary Fig. 5**). Therefore, none of the tested strains are devoid of TAS-R family unit that may suggest a functional role in the biology of the fly. These observations uncovered the fact that while the five longterm lab strains possess *TAS-3R*, four maintained *TAS-2R* and only one possess *TAS-X* (*Oregon R-C*). This could unveil that some TAS units are more critical than other to be maintained, while other can be lost without deleterious effect.

It has been suggested that *TAS-X* could be lost when strains are maintained for a while under laboratory standard conditions (Levis RW, unpublished data). If this is true, *Drosophila* strains harvested from the wild

and kept for decades are expected to have lost more frequently *TAS-X*, whereas freshly collected strains would still possess these sequences. Therefore, we first performed *in situ* hybridization on polytene chromosomes and/or PCR amplification on eleven unrelated strains isolated in diverse locations in the world and kept for several decades in our laboratory or in Stock Centers. The *in situ* hybridization results (**Supplementary Fig. 3**) summarized in **Table 1** shows that none of the strains are devoid of TAS loci as previously observed, and that the majority (8/11) are devoid of *TAS-X*. The three strains that still possess *TAS-X* are VAG_2, Wild_11C and BER_2 lines. BER_2 is the first strain found to carry *P* element in the 1950's (Anxolabehere, Kidwell et al. 1988). It is noteworthy that the Para Wirra strain tested in this study was previously described to contain a *P* element inserted at 1A, suggesting the presence of *TAS-X* at that time (Ronsseray, Lehmann et al. 1989), which were then lost under lab conditions.

To investigate the dynamics and the frequencies of their loss, we tested the status of *D. melanogaster* strains in their natural environment. Therefore, we sampled *Drosophila melanogaster* population between 2012 and 2015. In total we assayed 95 strains, among them 73 isofemale strains from Europe (45, including France and Austria), South America (27, including French Guiana and Brazil) and Africa (Ethiopia). Among this collection, polymorphism within three populations could be assessed, since 13 isofemale strains collected in Marsais (France), 19 in Niort (France) and 26 in Saul, an isolated town in the middle of the rainforest in French Guiana, were available. For the latter ones, strains were collected during tropical rain and dry seasons in case some environmental effects could be observed.

We first asked if some of the strains contain *TAS-X* using *in situ* hybridization on third instar larvae using the TAS probe. As shown in **Fig. 4** and **Supplementary Table 3** none of the recently collected strains were devoid of the three TAS loci, and that they all possess at least two TAS-R loci, as it was observed above for the strains kept for long period of time in laboratory conditions. Among strains isolated from the same location, we could observe polymorphism. For instance, the Marsais 53Ja strain bears the three TAS-R loci, whereas Marsais 80J and 87J have only *TAS-X* and *TAS-2R*. The striking results is that all the strains possess the *TAS-X* locus with either the *TAS-2R* or *TAS-3R* or both. Therefore, all strains tested recently collected from their wild environment have *TAS-X*. In order to confirm this observation, we performed PCR amplifications on these strains that are kept in two separate stocks, one raised at 18°C and one at 25°C. This was done at different time after the sampling. **Supplementary Table 3** summarized the PCR results for each of the strains tested. Consistent with *in situ* hybridization experiments, strains immediately tested after sampling showed PCR signals indicating that *TAS-X* repeats are always present on the tip of the *X* chromosome, except for one subline isolated in Ilha Grande, Rio de Janeiro (Brazil). For this latter, the strain was tested only three months post-sampling and after establishing two sublines one using one virgin female and a male (1♀x1♂) and one using 10 virgin females and 10 males (10♀x10♂). Since only the 1♀x1♂ subline appears to have lost the *TAS-X* and not the subline made of pooled individuals, it is tempting to speculate that the event of loss occurs after few generation in few individuals.

In order to test if the loss of *TAS-X* happened during maintenance of flies in the laboratory and, if possible, identify the dynamic of the loss, we retested after several months of maintenance in laboratory conditions at 18°C and 25°C. To our surprised, we were able to observed two strains collected in French Guiana (South America) that lost *TAS-X* one after 1 year at 25°C (Saul 13) and one after 18 months at 18°C (Saul A3). We decided to extend the analysis for the strains kept for two years in the lab: two of the strains isolated in Niort (Niort G2 and Niort P2.4) and kept at 18°C show loss of *TAS-X* while the sister strains kept at 25°C did not yet lose the locus. For the strains maintained at 25°C, five lost the locus (two in Marsais -53Ja and 90C-, two in Niort -P2.6 and L3- and one in Saul-FG24) while at 18°C the locus was maintained. A puzzling observation coming from all our analysis was that we never identified a strain with only one or not at all TAS loci, while terminal deletions are happening quite frequently. This observation seems to favor a function of the TAS loci. We do not know precisely yet if the loss involves environmental conditions that are apparently not the temperature (18°C vs 25°C), or simply due to the number of generations that favor loss of sequences by random telomeric breakage that grow non-essential in test tubes. Could this be a simple inbreeding effect???

During the course of our analysis, we observed that PCR amplifications products were sometimes less intense as stocks increasing number of generations after captured (**Supplementary Table 3**) giving us insight that number of *TAS-X* repeats within a strain could decline. An alternative hypothesis would be that some of the individuals in the pool of flies tested have lost the entire locus, as observed above. In parallel, while tested after its capture the Marsais 53Ja strain possessed all three TAS loci (on *X*, *2R*, *3R*), when performing *in situ* hybridization two years later on the subline reared at 25°C, however, *TAS-X* was missing. In order to confirm this observation, we reanalyzed by *in situ* hybridization ten male larvae (*i.e.* hemizygote for *X* chromosome) of this strain maintained at 18°C and ten male larvae maintained at 25°C. While the strain maintained at 18°C conserved the three TAS loci for all chromosomes tested (generation 23th; n=285), *TAS-X* locus was lost for the ten larvae tested on strain maintained at 25°C (generation 45th; n=297) (**Supplementary Fig. 4**). In order to increase the number of flies tested, we also individually PCR amplified 20 males from the subline maintained at 25°C (46th generation), and found that 18 over 20 actually have lost their *TAS-X* locus (**Supplementary Table 4**). Furthermore, for the next part, we performed double *in situ* hybridization using probes specific for TAS and *P* element to visualize if *P* element were inserted in TAS loci of different strains. For the strain maintained at 18°C, *P* hybridization was seen on *TAS-X* and *TAS-3R*, whereas the stain maintained at 25°C as expected had lost the hybridization signal on *TAS-X* but also lost the hybridization on *TAS-3R* accompanied by a decreasing in intensity of the *TAS-3R* signal (**Supplementary Fig. 4**). Therefore TAS loci, including *TAS-X* and probably *TAS-3R* are unstable sequences in the *Drosophila* genome.

In order to confirm the loss of *TAS-X* and reconstruct the chromosome ends after the deletion, we performed a series of PCR amplification using primers testing the presence of *TAS-D* and 0.9kb repeats, the two domains following the full length *TAS-B* repeat (**Fig. 1**). The reasons why we tested this structure was triggered by earlier *in silico* analysis performed on few sequences available in databases of the *X* tip: the minichromosome

containing the canonical *TAS-X* (Karpen and Spradling 1992), the *Canton-S* strain (Benos, Gatt et al. 2001) and *y^l; cn^l bw^l sp^l* reference genome (FlyBase (Adams, Celniker et al. 2000)). **Figure 1B** shows a schematic of the structure of *X* telomeres resulting when deletions occurred leaving behind the 0.9kb repeat (*Canton-S*) or not (*y^l; cn^l bw^l sp^l*) (**Fig. 1C**). This structure could be easily explained by random chromosomal breakage healed by transposition of a new telomeric TE. We have also revisited an extensive analysis studying subtelomeric regions on strains collected between 1990 and 2003 (Anderson, Song et al. 2008). Among these strains, none appears to possess *TAS-X* (*TAS-B* or *TAS-D*), while some possess the 0.9kb repeats, conferring them a *Canton-S* profile. We then assayed the *X*-telomeric structure of various strains including those who lost *TAS-X*. Using PCR primer sets specific for *TAS-B* repeat, *TAS-D* repeat and 0.9kb repeat, we first verified that the presence of *TAS-B* (positive signal on PCR amplification) was always accompanied by a *TAS-D* and a 0.9kb amplification. This was the case for all strains tested (**Supplementary Fig. 5A**). Using primers design for specifically amplify *TAS-D* subunit we optimized PCR conditions using *NA-P(1A)* as a positive control that lack *TAS-B* but still possess *TAS-D* (**Supplementary Fig. 3** and (Marin, Lehmann et al. 2000)), and as negative control *Canton-S* and *y^l; cn^l bw^l sp^l* (**Fig. 1B and C**, and **Supplementary Fig. 5B**). We also used PCR primers specifically directed against the 0.9kb repeats. *Canton-S* was used as positive control and *y^l; cn^l bw^l sp^l* as negative control (**Supplementary Fig. 5C**). It can be noted that the 0.9kb repeat present at the tip of the *X* chromosome in *Canton-S* strain is apparently stably maintained since 2001 (Benos, Gatt et al. 2001). Then, we tested several strains that were found to have a deletion of *TAS-X* in order to see if the deletion were progressively trimming the chromosomal end or if more drastic broken end occurs. **Supplementary Fig 5** exhibits a schematic of the results, where the majority of the recently collected strains that lost *TAS-X* (11/15) have lost the three domains tested (*TAS-B*, *TAS-D* and 0.9kb), where one retained the 0.9kb repeats and three still possess the *TAS-D* unit like the *NA-P(1A)*. The same bias toward the deletion of the three domains was observed when considering the strains maintained for decade in laboratory conditions. Therefore, it does not seem that the size of the loss correlate with the time spent in laboratory, favoring random terminal deletions hypothesis.

In conclusion, the results presented above indicate that the dynamic of the *TAS-X* seems to occur by terminal deletion. It is also possible that this locus would be able to gain repeats. Such events have been observed in strains that possess two *P-lacZ* transgenes inserted in *TAS-X* in the same orientation, have the same target site duplication and are inserted in a duplicated domain (Josse, Maurel-Zaffran et al. 2008). Here, we found that while the *TAS-X* are retained in the wild, in laboratory those sequences can be lost quite rapidly (two years) due to frequent chromosomal breakage. Therefore it might suggest that *TAS-X* is achieving a fundamental function in wild life that is not required in laboratory conditions.

3.3. *P* element in *TAS-X*

Previous studies revealed that some wild strains of *D. melanogaster* frequently carry *P* element inserted in *TAS-X* (Ajioka and Eanes 1989, Ronsseray, Lehmann et al. 1989). Functional evidence and molecular data (Ronsseray, Lehmann et al. 1991, Roche and Rio 1998, Marin, Lehmann et al. 2000, Josse, Teyssset et al. 2007) revealed that *P* is able to establish strong repression of euchromatic *P* elements, when a copy is inserted in *TAS-R* locus, mostly in *TAS-X*. It was later shown that *TAS-R* are piRNA clusters (Brennecke, Aravin et al. 2007) and that autorepression is essentially based on the piRNA biogenesis (Josse, Teyssset et al. 2007, Brennecke, Malone et al. 2008, Todeschini, Teyssset et al. 2010, de Vanssay, Bouge et al. 2012).

Hence, we reasoned that maintaining *TAS-X* could be due to several reasons including euchromatic *P* elements that have to be repressed by *P* inserted in *TAS*. This hypothesis would give a central role for *P* as well as other TE in the selection pressure for keeping *TAS* in wild population. Therefore, we performed double FISH experiment on polytene chromosomes of third instar larvae using *TAS-X* and *P* specific probes to test if colocalization can be observed. Control experiments were performed on *Lk-P*(1A) that contains two full length *P* elements (about 6 kb of homology with the probe) in *TAS-X* and *P-1152* (ca. 1.6 kb of homology with the probe) that contains two *P*-transgenes in *TAS-X*, *Harwich* that contains *P* elements on all chromosomal arms but no *TAS-X*, and *w¹¹¹⁸* devoid of *P* and *TAS-X* copies. From these control experiments, intensity of labelling could be correlated with size of the *P* sequence recognized by the probe (**Fig. 4, Supplementary Fig. 3, and Table 1**). We tested three strains from old stocks (*Harwich*, *Wild_5A* and *BER_2*) and nine strains recently collected from their natural environment. They were all bearing numbers of *P* elements on chromosomal arms (**Supplementary Fig. 3, Fig. 5, Supplementary Fig. 4, Table 1 and Supplementary Table 3**). This was anticipated as all strains are expected to be invaded by *P* sequences, since the last strain devoid of *P* was collected in 1970's in Ukraine (Anxolabehere, Kidwell et al. 1988). Among the 12 tested lines, *TAS-X* was present as described above (**Table 1 and Supplementary Table 3**). Colocalization between *P* element and *TAS* are observed on merging images for eight strains: two of them have a colocalization with *TAS-X* (*CacaoH1C* and *BER_2*), one with *TAS-2R* (*Kafoutine*), two with *TAS-3R* (*Chantemesle*, *Ilha Grande*), one with both *TAS-X* and *TAS-2R* (*Reth*), and one with both *TAS-X* and *TAS-3R* (*Marsais 53Ja*, *Marsais 86C*). While the presence of *P* elements on *TAS-X* was already documented in literature, no *P* elements on autosomal *TAS* were described, only few strains possessing *P*-transgenes (Josse, Maurel-Zaffran et al. 2008). The *BER_2* line is the first line identified bearing *P* elements. It shows strong *P* signal on *TAS-X* as well as several euchromatic *P*. This suggests that the insertion of *P* in *TAS-X* occurs during the course of genome invasion and a quick and strong repression was establish blocking any further transposition. This was confirmed by PCR amplification identifying only full-length *P* elements (data not shown). The Brazilian subline that happened to be the only clear line devoid of *TAS-X* after being collected appears to have a *P* inserted in *TAS-3R*. Telomeric *3R P* insertions could repress efficiently euchromatic *P* elements, as it has been shown for *P*-transgene inserted in *TAS-3R* (Dufourt, Dennis et al. 2014). As a result, it could soften pressure to keep *TAS-X* sequences. From

these results, *P* element in *TAS-X* could be considered somehow one of the player in this phenomenon. However four lines do not present colocalization between *P* and *TAS* locus indicating that *P* is not the unique motor to maintain pressure of selection. Moreover *TAS* loci were present before the *P* element invasion, reflecting the fact that *P* could not be the unique factor favoring the *TAS* maintenance. One can also speculate that other family of TE could insert in *TAS* as well and have used these loci to establish their own repression. But this could not account for the fact that *TAS-X* can be lost so easily in laboratory condition which suggests that this locus appears less required in lab for the establishment of repression. For instance *P* element inserted in secondary piRNA cluster inefficient in natural environment would awake and start to produce efficiently piRNA in laboratory conditions. Alternatively, *TAS-X* would be required for other functions wild life and not in lab, like for instance celerity, search for food, or looking for partners.

3.4. *TAS* in *Drosophila melanogaster* sister species

After the structural identification of *TAS* locus and identification of the *TLL* domain, we asked about their status and their evolutionary origin in *D. melanogaster* closely related species. We performed BLAST searches using domain F sequence against other *Drosophila* species. We found several scaffolds in *D. sechellia* genome and to some extent in *D. simulans* genome. In *D. sechellia*, at least nine scaffolds presented *ca.* 80% of identity with the *TLL* domain of *D. melanogaster*. We decided to focus on one of them, Scaffold 31, because it seemed to contain several *TLL* domains. The analysis allowed us to uncover that it is composed of at least five repeats of 855bp. Each repeat contains one *TLL* sequence of 191bp, an unrelated internal sequence of 403bp, another *TLL* of 138 bp followed by another unrelated sequence of 123bp (**Fig. 5A**). This scaffold bears another 138 bp repeat, flanking a region containing one domain G (>90% identity) at 8.8kb from the *GM16406* gene, ortholog of *D. melanogaster* *Map205* the most distal gene before the *TAS-3R* and on the other end harbors a *TART* telomeric element (**Fig. 5A**). Another scaffold, Scaffold 14, appears also of interest because it contains two copies of the internal sequence, two full length copies of the 855bp repeat, two subregions corresponding to domain G and the ortholog of *l(2)gl* gene, the most distal gene before the *TAS-2L* in *D. melanogaster*. These results are of interest because they suggested that the repeats identified could be telomeric. Using the *D. sechellia* repeat of 855 bp, BLAST searches identified within the *D. simulans* genome eight scaffolds, six of which restricted to the internal sequence. Within Scaffold 3R, a domain of 403bp containing two *TLL* (>80% identity) was found, followed by one domain G (90% identity), one subregion of domain C (86 % identity) and at 2.3kb the *GD16184* gene, the ortholog of *D. melanogaster* *Map205* gene (**Fig. 5B**). In *D. erecta*, seven scaffolds were identified containing several *TLL* repeats of smaller fragments (60-65 bp) with 85% of identity with the 855bp repeat. Among them, Scaffold 4820 contains four tandemly repeated domains of 998bp including an 183bp and 142bp matching with the two *TLL* of *D. sechellia*. Thirteen kb further down, the *GG24690* gene ortholog of *l(2)gl* of *D. melanogaster* is found. Finally, in *D. yakuba*, six scaffolds containing

about 170bp of the internal sequence are found and one scaffold, scaffold 2L, that contains two *TLL* of 60 bp spaced by a region of 104bp (85-90 % identity) were identified 2kb away from the *GE16527* gene the ortholog of *l(2)gl*. We did not find *TLL*-like sequence in others species. This analysis revealed that *TLL*-like sequences in sister species exist, are repeated and their conserved syntheny seems to indicate that they could be telomeric. We then performed *in situ* hybridization on polytene chromosomes using a probe prepared with the 855 bp repeat of *D. sechellia*. **Figure 5D** shows that *TLL*-like sequences are indeed telomeric in *D. sechellia*. Three telomeres were hybridized with the probe. No hybridization was found on chromosomal arms, neither on chromocenter. We used the same probe to test the *D. simulans* polytene chromosomes and identified at least two telomeric signals. Since *TLL* of *D. sechellia* is close in term of sequence to *D. melanogaster TLL*, we used the same probe on polytene chromosomes of *D. melanogaster*. On top of a telomeric *TAS-X* hybridization, the 855 bp probe of *D. sechellia* strongly labelled the *TAS-2L* and *TAS-3L* repeats. A faint labeling on *TAS-3R* may be also observed. BLAST searches using the 855bp of *D. sechellia* identified a 178bp sequence (89%) at 1kb of the *TAS-3R* repeats. This sequence corresponds to the 191bp-*TLL* repeat of *D. sechellia* that could be recognized by the probe. Chromocenter hybridizations were also detected that could be explained by numbers of BLAST hits found on *Y* chromosome, *X* and *2R* centromeres corresponding to the 526bp internal sequence as well as several hits with the *TLL*. Therefore using *D. melanogaster* *TAS* sequences, we were able to identify the orthologs in the *D. sechellia*, *D. simulans* and *D. erecta* genome. Like in the *D. melanogaster* genome, these loci are telomeric supporting the assumption that the function of *TAS*, if any, could be as well required in these genomes.

3.5. *TAS* are piRNA producer loci

In *D. melanogaster*, the *TAS* loci have been shown to be piRNA producer loci (Brennecke, Aravin et al. 2007). Therefore we asked if the sequences identified above in *D. sechellia* and *D. simulans* could also be part of piRNA producer locus. To address this question, we sequenced small RNA from ovaries of these *Drosophila* species, and screened for piRNA production. We first characterized piRNA production of *TAS* loci from the *D. melanogaster* genome in different strains used in this study. The profiles of piRNA are shown on (**Supplementary fig. 6**) when considering the full length *TAS-X* in *P-1152* and *Oregon-R-C* strains that contains *TAS-X*. The graphs show that small RNA of the size of piRNA (23-29 nt) are produced in ovaries. These piRNA are produced throughout the sequence, with a large proportion coming from the *INV-4* solo LTR that are also present on other piRNA loci (for instance 42AB (Iwasaki, Siomi et al. 2015)). This can be visualized with the graphs obtained for the *y¹*; *cn¹ bw¹ sp¹* and *CantonS* that are devoid of *TAS-X* locus but possess the 42AB cluster (**Supplementary Fig. 6**). Interestingly, the *y¹*; *cn¹ bw¹ sp¹* shows piRNA matching to its 3' region corresponding to the domain C that is present on the *TAS-2R* (**Fig. 2**), suggesting that *TAS-2R* present in this genome is also producing piRNA. Therefore, we restricted this analysis to the domain F unique

on *TAS-X*. The graph shows that piRNA from domain F are produced from strains harboring *TAS-X* repeats, *i.e.* *Oregon-R-C* (devoid of *P* sequences), *Lk-P(1A)* and *P-1152*, containing *P* and *P*-transgene respectively (**Fig. 6 and Supplementary Fig. 6**). No piRNA corresponding to domain F was identified in strains devoid of *TAS-X* (*y¹*; *cn¹* *bw¹* *sp¹* and *CantonS*). We also analyzed the production of piRNA from strains recently collected; Chantemesle, Reth (French metropole) and Cacao H1C (Saul, French Guiana) that were characterized above to contain *TAS-X*. A polymorphism can be observed for the production of piRNA among the three strains: the same profile containing the same discrete pics are observed but while Chantemesle is producing comparable amount, the Reth and Cacao H1C strains are producing less piRNA corresponding to domain F than strains maintained for decades in laboratory, *i.e.* *Lk-P(1A)*, *P-1152* and *Oregon-R-C*, (**Fig. 6 and Supplementary Fig. 6**). Therefore, when *TAS-X* is present in a strain, it is always producing piRNA, and the presence of an inserted *P* element (*Lk-P(1A)*, Reth, Cacao H1C) or *P*-transgene (*P-1152*) in *TAS-X* does not affect the production of the locus when compare to a strain devoid of *P* element in *TAS-X* (*Oregon-R-C*, Chantemesle). We then examine the profile of the 0.9kb repeats in the seven strains sequenced (*Lk-P(1A)*, Cacao H1C, Chantemesle, Reth, *P-1152*, *Oregon-R-C* and *Canton-S*). **Fig. 6 and Supplementary Fig. 6** show that the 0.9kb repeats are also producing piRNA. This production appears independent of the TAS repeats as *Canton-S* strain is producing whereas it lacks *TAS-X* locus. We can notice that *Lk-P(1A)* produce a high level of a 23 nt small RNA that appears to misrepresent the piRNA profile. Ultimately we also analysed the TAS-L piRNA profile. This locus is also producing piRNA in all of the strains, except the *Lk-P(1A)* strain which appears to bear an inefficient locus for a reason that we could not explain (**Fig. 6 and Supplementary Fig. 6**). The size distribution of piRNA produced by domain F, TAS-L and 0.9kb locus shows that these loci are dual strand piRNA clusters. The origin of transcription of *TAS-X*, represented by the domain F, is still unknown. We could hypothesize that sense piRNA could be produced from a promoter located in the vicinity of the telomeric elements. The production of antisense piRNA is inefficient compare to the sense piRNA production suggesting that *INV4* soloLTR are weak promoters. This imbalance between the convergent promoters results in an asymmetric dual-strand piRNA clusters.

To confirm the localization of *P* element within TAS locus observed by *in situ* hybridization on polytene chromosome, we also looked and found numbers of piRNA homologous to *P* element. *P* element piRNA are produced all along the sequence in *Lk-P(1A)* as previously published (Brennecke, Malone et al. 2008). The production of *P* piRNA shown in **Fig. 6** by the Cacao H1C, Reth and Chantemesle strains correlates with the colocalization of *P* within *TAS* locus identified above (**Fig. 4**). A functional test revealed that those piRNA are functional since they are able to repress Gonadal Dystrophy (GD), a syndrome provoked by unregulated euchromatic *P* elements (**data not shown**). The profile of *P* piRNA from Chantemesle and Reth resembles to the well described form of internally deleted *P* element, called KP, shown to be able to repress efficiently euchromatic *P* (Black, Jackson et al. 1987). Here we show that the germline repression is most likely due to piRNA production. Taken altogether, Brennecke et al (Brennecke, Aravin et al. 2007) already described TAS

as piRNA producer loci. Here we show the precise profile of these dual-strand piRNA clusters: the *TAS-X* represented by the domain F, the 0.9kb repeat as independently expressed from the *TAS-X* unit (see *Canton-S*) and the *TAS-L* family. It should be noted that no significant differences are observed for a specific locus in different genetic background including long time established strains (y^1 ; cn^1 bw^1 sp^1 , *Canton-S*, *Oregon-R-C*, *LK-P(1A)* and *P-1152*) in laboratory vs strains isolated recently from the wild (Chantemesle, Reth, CacaoH1C). However, this apparent homogeneity for piRNA cluster is not seen for profile of TE, like *P* element, simply because they can exist as full length (*LK-P(1A)*, Cacao H1C) or as deleted copies (Chantemesle) and inserted in different piRNA cluster: Reth is the result of insertion in at least two piRNA clusters (*TAS-X* and *TAS-2R*), Chantemesle has an insertion in *TAS-3R* and CacaoH1C and *LK-P(1A)* have at least an insertion in *TAS-X* showing (Supplementary Table 3).

As identified in the above section, *D. simulans* and *D. sechellia* possess sequences homologous to *TAS-R* that are telomeric and repeated. Hence, we reasoned that those sequences could have the capacity to produced piRNA as well. We analyzed small RNA extracted from ovaries of *D. melanogaster* sister species and looked for piRNA homologous to the repeats identified in each species. For *D. sechellia*, the 855bp repeats appear to be a region that is capable of producing piRNA (23-29nt), as well as for the repeats found in *D. simulans* and to the *TLL* domain that matches in *D. melanogaster* (Fig. 5). As for the *D. melanogaster TLL*, the *TLL* domains found in *D. sechellia* and *D. simulans* are also part of a subtelomeric region producing piRNA (Supplementary Fig. 7). Therefore, we unveiled that the *TAS* repeats identified in the three *Drosophila* species have *TLL* for origin, are repeated and are dual-strand piRNA producer loci, as piRNA on both sense are produced.

3.6. Evolution of subtelomeric piRNA clusters and concluding remarks

By our analysis, we uncovered that members of the *TAS-R* family are more related to each other than originally anticipated. While the *INV-4* LTR were known to be the factors defining the family, we found that other relatively conserved domains are also participating in the structure of the family. The conservation of these domains could reflect function as for instance factors involved in the expression of these piRNA clusters in germline. Actually a number of factors (*i.e* PC, E(Z), H3Me3K27) have been found enriched on *TAS-R* (Boivin, Gally et al. 2003, Andreyeva, Belyaeva et al. 2005, Antao, Mason et al. 2012). Through a deep analysis of domain F, we found it located only on *TAS-X*, and containing a region of 158 bp that we called *TLL*. *In silico* analyses using *TLL* identified that *TAS-R* family is related to the *TAS-L* family repeats and to repeated subtelomeric regions in only *D. sechellia*, *D. simulans*, *D. yakuba* and *D. erecta*. Therefore *TLL* appears restricted to the melanogaster subgroup. In agreement with the sequence that we found, we could propose that *TLL* was duplicated in the common ancestor of the subgroup that could then be lost or evolved along with the speciation of the *Drosophila*.

Interestingly, it was not possible to find an origin for the different domains constituting the TAS loci, except for the *INV4* solo LTR. It is tempting to speculate that TAS has derived from ancient horizontal transfers of TE that get inserted into the TAS locus, as the *P* element does recently. Once inserted, piRNA would be produced that will efficiently repress euchromatic TE that could then degenerate. Therefore, TAS will be made of fragments of extinct TE. TAS have been proposed to be involved in the biology of telomeres (Marzullo and Gatti 2016). Nevertheless, subtelomeric repeats on the *X* chromosome appears to be able to expand or diminish their number of repeats quite rapidly by few chromosomal breakage leading to a complete deletion of the locus or by unequal cross-over. Therefore, we were puzzled by the fact that the *TAS-X* seems required for flies living in their natural environment. This might reflect a more complex requirement of *TAS-X* that might include regulation of genes dispensable for *in vitro* conditions. Therefore domestication of *Drosophila* seems to alleviate pressure on some genes. A recent publication identified that such domestication might have influence on neurogenesis genes (Stanley and Kulathinal 2016). Further experiments will be of course required to test the relation between domestication and subtelomeric repeats.

4. Availability:

Accession numbers

Acknowledgements

We thank Clément Carré, Emile Civel, Jean-Michel Gibert, Aurélie Hua-Van, Claude Lespinet, Pauline Marie, Valerie Ribeiro, Michèle Teyssset, Anne Laure Todeschini, and Cristina Vieira for their help in collecting flies, the Bloomington Stock Center for providing stocks, FlyBase.org for providing valuable databases and ARTbio/Bioinformatics analyses platform at the IBPS for the development of valuable tools. We thank Jean-François Gilles and Susanne Bolte for the help with confocal microscopy. Confocal work was carried out at the Institute of Biology Paris-Seine Imaging Facility that is strongly supported by the “Conseil Regional Ile-de France”, the French national research council (CNRS) and Sorbonne University, UPMC Univ Paris 06. We thank Ritha Zamy for technical assistance, Jean-Michel Gibert for critical reading of the manuscript. This work was supported by fellowships from the Ministère de l’Enseignement Supérieur et de la Recherche to A.A-L and by grants from the Association de la Recherche contre le Cancer (Fondation ARC, SFI20121205921, SFI20131200470), the Fondation pour la Recherche Médicale, (FRM DEP20131128532), the Association Nationale de la Recherche (ANR, project “platisipi”) to S.R., and the University Pierre et Marie Curie (Emergence EME1223) to L.T. Sampling of *D. melanogaster* in French Guiana was funded by a CEBA grant (MONISPEC ANR-10-LABX-25-01) to W.J.M. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript. The authors declare that no competing interest exists.

Figure legends

Figure 1. Subtelomeric organization of the X chromosome from different strains of *D. melanogaster*

A. The subtelomeric region of *Dp1187* minichromosome is composed by *HeT-A* telomeric element followed by four TAS repeats: *TAS-A*, *TAS-B*, *TAS-C* and *TAS-D*. Two repeats of 0.9kb are also found followed by the euchromatic region. This clone does not include the first gene. B. *Canton-S* strain lacks the *TAS-X* repeats but has the 0.9kb repeats. *CG17636* is the first gene found on the X chromosome. C. *y¹; cn¹ bw¹ sp¹* lacks all the subtelomeric region including *TAS-X* repeats as well as the 0.9 kb repeats.

Figure 2. Schematic representations of the TAS-R family and the TLL domains.

TAS-R family is composed by the *TAS-2R* (A), *TAS-3R* (B) and *TAS-X* (C). They are constituted by several domains organized in a specific order for each locus. These domains have *INV-4* sequences corresponding to solo LTR (amber), domain C (pink), domain D (yellow), domain E (blue), domain F (green) and domain G (violet). For a better comparison, the *INV-4* solo LTR are at left side and telomere specific repeats for each telomere at right. The size of each type of repeat is indicated. *TLL* (red boxes) can be found in the domain F of the *TAS-X* (C), in the 0.9 kb repeat found in the X chromosome (D) and in the TAS-L (*TAS-2L* and *TAS-3L*) (E). The TAS-L appears to be made of several *TLL* repeats. The *TLL* of the F domain is taken as reference here for the values of percentage of identity between sequences. Identity between the 169bp of 0.9kb repeat and the 149 bp found in *TAS-2L* is 85%. Sizes of sequences (bp) are indicated below the red boxes.

Figure 3. P elements and TAS-R repeats on polytene chromosomes of laboratory strains.

Polytene chromosomes of third instar larvae were hybridized with specific probes for TAS-R repeats (first column, red) and P-elements (second column, green). DAPI staining shows chromosomes in blue. TAS-R sequences are present on telomeres of the X and 2R chromosomes in *Lk-P(1A)* strain and of the X, 2R and 3R chromosomes in *Oregon R-C* strains. As expected P element is present on the X telomere of the *Lk-P(1A)* and *Oregon R-C* is devoid. The merge images on third and fourth columns clearly show colocalization (white star) on X chromosome for the *Lk-P(1A)* strain. NB: Aspecific labellings can be sometimes seen that correspond to nuclear membranes, like the one visible with the P probe in the *Lk-P(1A)* chromosomes.

Figure 4. Localization of TAS-R repeats and P elements in strains recently collected from their natural environment. The first column shows the hybridization obtained with the TAS-R probe (green), the second

column was obtained using a full length *P* element (red) showing numerous *P* elements on chromosomal arms. On the third column, chromosomes are DAPI stained allowing to visualize colocalization of *P* element on TAS locus for the Ilha Grande, Cacao H1C, Chantemesle and a double colocalization for the Reth strains. Marks (white) show the labelling of the probe to telomeres. The white arrowheads indicate the chromosomes labelled by the TAS probes. Colocalizations are shown with a star.

Figure 5. TLL in *D. melanogaster* sister species

A. Schematic representation of *D. sechellia* Scaffold 31. Orange boxes are *D. sechellia* repeats of 855bp. One repeat is composed of two subrepeats (138 and 191 bp), having homology with *TLL* sequence of *D. melanogaster*. *GM16406* gene is found about 22800 bp upstream of the 855bp repeats. This gene is the *D. melanogaster* ortholog of *Map205* the last gene before *TAS-3R*. Hatched boxes correspond to longer *D. sechellia* repeats (1143bp). The purple and pink boxes are the sequences corresponding to the *D. melanogaster* DomC and DomG. The last sequence in this scaffold is a telomeric *TART* element. **B.** Schematic representation of *D. simulans* Scaffold 3R. Orange boxes are 855bp like sequences, containing one sequence of 202bp and 59bp sequence corresponding to the TLL of *D. sechellia*. The 142 bp interspersed sequence is specific of the scaffold 3R. The purple and pink boxes are the sequences corresponding to the *D. melanogaster* DomC and DomG. *GD16184* gene is found about 2kb up stream of the DomC and is the ortholog of *D. melanogaster* *Map205* gene, the last gene before the *TAS-3R*. **C.** *In situ* hybridization on polytene chromosome of *D. sechellia*, *D. simulans* and *D. melanogaster* (Oregon R-C). The probe was made by amplification of the *D. sechellia* repeats of 855nt (red). Marks (white) show the labelling of the probe to telomeres **D-E** Readmaps and size distributions of ovarian small RNA. Positive and negative reads are sense and antisense reads respectively. **D.** Readmaps show the abundance of 23 to 29 nt small RNA matching to one *D. sechellia* repeat (855bp) (E1), *D. simulans* (E2) and to *D. melanogaster* (E3) TLL sequence. piRNA are produced from the whole length of the 855bp repeat (E1). The *D. simulans* TLL repeats appears to be a poor piRNA producer at least in this strain (E2). Discrete pics of piRNA production could be observed for the DomF TLL domain (E3). **E.** Size distributions histograms show the length of small RNA matching to each reference sequence *D. sechellia* repeat (855bp) (F1), *D. simulans* (F2) and *D. melanogaster* TLL sequence (F3).

Figure 6. piRNA production for recently collected *Drosophila* strains in the wild.

Readmaps show the abundance of 23 to 29 nt small RNA matching to each reference sequence: the domain F which is specific of *TAS-X*, the 0.9kb, the *TAS-L* and the *P* element. Size distributions histograms show the length of small RNA matching to each reference sequence (Domain F, *P*-element, 0.9kb, *TAS-2L*). Positive and negative reads are sense and antisense reads respectively. *Lk-P(1A)* is a positive control since this strain

649 contains two *P*-elements in the *TAS-X*. The three recent wild strains (Chantemesle, Reth and Cacao H1C)
650 produce piRNA (23-29nt) for the Domain F and have the similar profile as in *LK-P(1A)*. These strains also
651 produce piRNA of *P* element (23-29nt) but have different profiles. Reth and Cacao show piRNA for the whole
652 length of the *P* element sequence which is not the case for Chantemesle. Also the level of piRNA produced is
653 different than in *Lk-P(1A)* which has two *P* elements reflecting the large number of copies present in the
654 recently collected strains. For the 0.9kb repeat and the *TAS-L* the three wild strains have the same profile for
655 piRNA production but not *Lk-P(1A)*.

656

Supplementary Figure 1. Muscle alignment of the 4 *TAS-X* repeats (*TAS-A*, *TAS-B*, *TAS-C*, and *TAS-D*) from *Dp1187* minichromosome (see Fig.1A). The domain C (pink), domain D (yellow), domain E (blue), domain F (green) and *INV-4* solo LTR (black box) are shown. The *TLL* domain is depicted by a red bar under the sequence alignment. Primers used in this study are shown, for more details see the Materials and Methods section.

Supplementary Figure 2. Comparison of the *INV-4* sequences from the canonical retrotransposon and subrepeats of the *TAS-R*. (A) Representation of the *INV-4* from the canonical element. (B) Representation of the *INV-4* subrepeats identified on the *TAS-B* sequence, the reference repeats for the *TAS-X* (see Supplementary Fig.1) and on the *TAS-2R* (C) and *3R* (D) respectively. Numbers indicate the coordinates on the corresponding sequences and numbers in parenthesis indicate the size of the domains. On this representation, the 172 bp central cores common between the three *TAS-R* loci and the canonical *INV-4* TE are shown, as well as the 3' and 5' extensions on each side of the *TAS-R* loci.

Supplementary Figure 3. Localization of *P* elements and *TAS-R* repeats on polytene chromosomes on various strains.

Polytene chromosomes of third instar larvae were hybridized with probes specific for *TAS-R* repeats (first column, red) and *P*-elements (second column, green). *TAS-R* sequences can be observed on subtelomeric regions of *TAS-X*, *2R* and *3R*. Whole chromosomes are shown on merge images on the third column. Telomeric magnifications display on fourth column colocalization between *P* element and *TAS* locus for *P-1152* and *BER_2* strains. No *P*-element are observed inserted into other telomeres or into telomere of other strains. A great number of copies are present in chromosome arms for Harwich and *BER_2*. Less *P*-element can be detected for Wild_5A. The white arrows indicate the chromosomes labelled by the *TAS* probes. Colocalizations are shown with a star.

Supplementary Figure 4. Loss of *TAS-X* in a 2 years period in laboratory conditions.

In situ hybridization on polytene chromosomes of the Marsais 53Ja strains maintained at 18°C (A) and 25°C (B) using *TAS* and *P* element specific probes. A. At 18°C, the three *TAS-R* loci are labelled with the *TAS* probe (first column). *TAS-X* and *TAS-3R* present colocalization of the *P* element. NB: the two homologous chromosomes of the *3R* arm have been separated at the chromosomal end during the squash, showing the labelling on each of the homologous chromosome. B. At 25°C, only two *TAS* (*2R* and *3R*) are labelled suggesting an apparent loss of *TAS-X* at 25°C. Note that the labeling of *P* element is lost at the *TAS-X* as well as

at the *TAS-3R*. This suggests a loss of some of the *TAS-3R* repeats that contained the insertion of the *P* element that can be correlated with a fainter *TAS-3R* band. n, shows the number of nuclei showing labeling. The white arrows indicate the chromosomes labelled. Colocalizations are shown with a star.

Supplementary Figure 5. Schematic representation of *X* telomeric ends of several strains kept for decades in laboratory (left column) or recently collected from wild (right column)

By PCR amplification, using specific primers for *TAS-B*, *TAS-D* and 0.9kb repeats, we were able to identify the telomeric structure of tested strains.

Supplementary Figure 6. TAS loci piRNA production in *Drosophila*

Graphs show the abundance and size distributions of the 23 to 29 nt small RNA matching to each reference sequence: full length *TAS-X*, domain F which is specific of *TAS-X*, 0.9kb and *TAS-L*. Positive and negative values are sense and antisense reads respectively. N/D, Not Determined, when no read corresponding to the reference sequence are found.

Supplementary Fig. 7. *Drosophila sechellia* and *Drosophila simulans* subtelomeric regions are piRNA clusters.

Readmaps show the abundance of 23 to 29 nt small RNA matching to each reference sequence subtelomeric region containing the *TLL* domains identified in *D. sechellia* and *D. simulans* up to the *Map205* ortholog gene (non included).

711 References

- 712 Abad, J. P., B. De Pablos, K. Osoegawa, P. J. De Jong, A. Martin-Gallardo and A. Villasante (2004).
 713 "Genomic analysis of *Drosophila melanogaster* telomeres: full-length copies of HeT-A and TART elements
 714 at telomeres." *Mol Biol Evol* 21(9): 1613-1619.
- 715 Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W.
 716 Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G.
 717 Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y. H. Rogers, R. G. Blazej, M.
 718 Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G. Baxter, G. Helt, C. R. Nelson, G. L. Gabor, J. F. Abril, A.
 719 Agbayani, H. J. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, L.
 720 Bayraktaroglu, E. M. Beasley, K. Y. Beeson, P. V. Benos, B. P. Berman, D. Bhandari, S. Bolshakov, D.
 721 Borkova, M. R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K. C. Burtis, D. A. Busam, H. Butler, E. Cadieu,
 722 A. Center, I. Chandra, J. M. Cherry, S. Cawley, C. Dahlke, L. B. Davenport, P. Davies, B. de Pablos, A.
 723 Delcher, Z. Deng, A. D. Mays, I. Dew, S. M. Dietz, K. Dodson, L. E. Doup, M. Downes, S. Dugan-Rocha, B. C.
 724 Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E.
 725 Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M.
 726 Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J.
 727 Howland, M. H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G. H. Karpen, Z. Ke, J. A. Kennison, K. A. Ketchum,
 728 B. E. Kimmel, C. D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A. A. Levitsky, J. Li, Z. Li, Y.
 729 Liang, X. Lin, X. Liu, B. Mattei, T. C. McIntosh, M. P. McLeod, D. McPherson, G. Merkulov, N. V. Milshina,
 730 C. Mobarry, J. Morris, A. Moshrefi, S. M. Mount, M. Moy, B. Murphy, L. Murphy, D. M. Muzny, D. L.
 731 Nelson, D. R. Nelson, K. A. Nelson, K. Nixon, D. R. Nusskern, J. M. Pacleb, M. Palazzolo, G. S. Pittman, S.
 732 Pan, J. Pollard, V. Puri, M. G. Reese, K. Reinert, K. Remington, R. D. Saunders, F. Scheeler, H. Shen, B. C.
 733 Shue, I. Siden-Kiamos, M. Simpson, M. P. Skupski, T. Smith, E. Spier, A. C. Spradling, M. Stapleton, R.
 734 Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A. H. Wang, X. Wang, Z. Y. Wang, D. A.
 735 Wassarman, G. M. Weinstock, J. Weissenbach, S. M. Williams, WoodageT, K. C. Worley, D. Wu, S. Yang,
 736 Q. A. Yao, J. Ye, R. F. Yeh, J. S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X. H. Zheng, F. N. Zhong, W.
 737 Zhong, X. Zhou, S. Zhu, X. Zhu, H. O. Smith, R. A. Gibbs, E. W. Myers, G. M. Rubin and J. C. Venter (2000).
 738 "The genome sequence of *Drosophila melanogaster*." *Science* 287(5461): 2185-2195.
- 739 Ajioka, J. W. and W. F. Eanes (1989). "The accumulation of P-elements on the tip of the X chromosome in
 740 populations of *Drosophila melanogaster*." *Genet Res* 53(1): 1-6.
- 741 Anderson, J. A., Y. S. Song and C. H. Langley (2008). "Molecular population genetics of *Drosophila*
 742 subtelomeric DNA." *Genetics* 178(1): 477-487.
- 743 Andreyeva, E. N., E. S. Belyaeva, V. F. Semeshin, G. V. Pokholkova and I. F. Zhimulev (2005). "Three
 744 distinct chromatin domains in telomere ends of polytene chromosomes in *Drosophila melanogaster* Tel
 745 mutants." *J Cell Sci* 118(Pt 23): 5465-5477.
- 746 Antao, J. M., J. M. Mason, J. DeJardin and R. E. Kingston (2012). "Protein landscape at *Drosophila*
 747 *melanogaster* telomere-associated sequence repeats." *Mol Cell Biol* 32(12): 2170-2182.
- 748 Anxolabehere, D., M. G. Kidwell and G. Periquet (1988). "Molecular characteristics of diverse populations
 749 are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile P
 750 elements." *Mol Biol Evol* 5(3): 252-269.
- 751 Benos, P. V., M. K. Gatt, L. Murphy, D. Harris, B. Barrell, C. Ferraz, S. Vidal, C. Brun, J. Demaille, E. Cadieu,
 752 S. Dreano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, D. Borkova, B. Minana, F. C. Kafatos, S. Bolshakov,
 753 I. Siden-Kiamos, G. Papagiannakis, L. Spanos, C. Louis, E. Madueno, B. de Pablos, J. Modolell, A. Peter, P.
 754 Schottler, M. Werner, F. Mourkioti, N. Beinert, G. Dowe, U. Schafer, H. Jackle, A. Bucheton, D. Callister, L.
 755 Campbell, N. S. Henderson, P. J. McMillan, C. Salles, E. Tait, P. Valenti, R. D. Saunders, A. Billaud, L.
 756 Pachter, D. M. Glover and M. Ashburner (2001). "From first base: the sequence of the tip of the X

757 chromosome of *Drosophila melanogaster*, a comparison of two sequencing strategies." Genome Res
758 11(5): 710-730.

759 Biessmann, H. and J. M. Mason (2003). "Telomerase-independent mechanisms of telomere elongation."
760 Cell Mol Life Sci 60(11): 2325-2333.

761 Black, D. M., M. S. Jackson, M. G. Kidwell and G. A. Dover (1987). "KP elements repress P-induced hybrid
762 dysgenesis in *Drosophila melanogaster*." EMBO J 6(13): 4125-4135.

763 Boivin, A., C. Gally, S. Netter, D. Anxolabehere and S. Ronsseray (2003). "Telomeric associated sequences
764 of *Drosophila* recruit polycomb-group proteins in vivo and can induce pairing-sensitive repression."
765 Genetics 164(1): 195-208.

766 Boussy, I. A. and M. Itoh (2004). "Wanderings of hobo: a transposon in *Drosophila melanogaster* and its
767 close relatives." Genetica 120(1-3): 125-136.

768 Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam and G. J. Hannon (2007).
769 "Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*." Cell
770 128(6): 1089-1103.

771 Brennecke, J., C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark and G. J. Hannon (2008). "An
772 epigenetic role for maternally inherited piRNAs in transposon silencing." Science 322(5906): 1387-1392.

773 de Vanssay, A., A. L. Bouge, A. Boivin, C. Hermant, L. Teyssset, V. Delmarre, C. Antoniewski and S.
774 Ronsseray (2012). "Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus." Nature
775 490(7418): 112-115.

776 Dufourt, J., C. Dennis, A. Boivin, N. Gueguen, E. Theron, C. Goriaux, P. Pouchin, S. Ronsseray, E. Brassat
777 and C. Vaury (2014). "Spatio-temporal requirements for transposable element piRNA-mediated silencing
778 during *Drosophila* oogenesis." Nucleic Acids Res 42(4): 2512-2524.

779 Gouy, M., S. Guindon and O. Gascuel (2010). "SeaView version 4: A multiplatform graphical user
780 interface for sequence alignment and phylogenetic tree building." Mol Biol Evol 27(2): 221-224.

781 Hermant, C., A. Boivin, L. Teyssset, V. Delmarre, A. Asif-Laidin, M. van den Beek, C. Antoniewski and S.
782 Ronsseray (2015). "Paramutation in *Drosophila* Requires Both Nuclear and Cytoplasmic Actors of the
783 piRNA Pathway and Induces Cis-spreading of piRNA Production." Genetics 201(4): 1381-1396.

784 Iwasaki, Y. W., M. C. Siomi and H. Siomi (2015). "PIWI-Interacting RNA: Its Biogenesis and Functions."
785 Annu Rev Biochem 84: 405-433.

786 Josse, T., C. Maurel-Zaffran, A. de Vanssay, L. Teyssset, A. L. Todeschini, V. Delmarre, N. Chaminade, D.
787 Anxolabehere and S. Ronsseray (2008). "Telomeric trans-silencing in *Drosophila melanogaster*: tissue
788 specificity, development and functional interactions between non-homologous telomeres." PLoS One
789 3(9): e3249.

790 Josse, T., L. Teyssset, A. L. Todeschini, C. M. Sidor, D. Anxolabehere and S. Ronsseray (2007). "Telomeric
791 trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation." PLoS
792 Genet 3(9): 1633-1643.

793 Kaminker, J. S., C. M. Bergman, B. Kronmiller, J. Carlson, R. Svirskas, S. Patel, E. Frise, D. A. Wheeler, S. E.
794 Lewis, G. M. Rubin, M. Ashburner and S. E. Celniker (2002). "The transposable elements of the
795 *Drosophila melanogaster* euchromatin: a genomics perspective." Genome Biol 3(12): RESEARCH0084.

796 Karpen, G. H. and A. C. Spradling (1992). "Analysis of subtelomeric heterochromatin in the *Drosophila*
797 minichromosome Dp1187 by single P element insertional mutagenesis." Genetics 132(3): 737-753.

798 Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of
799 short DNA sequences to the human genome." Genome Biol 10(3): R25.

800 Levis, R. W. (1989). "Viable deletions of a telomere from a *Drosophila* chromosome." Cell 58(4): 791-801.

801 Levis, R. W., R. Ganesan, K. Houtchens, L. A. Tolar and F. M. Sheen (1993). "Transposons in place of
802 telomeric repeats at a *Drosophila* telomere." Cell 75(6): 1083-1093.

Marin, L., M. Lehmann, D. Nouaud, H. Izaabel, D. Anxolabehere and S. Ronsseray (2000). "P-Element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric P copy." Genetics 155(4): 1841-1854.

Marzullo, M. and M. Gatti (2016). "Telomere fusion in *Drosophila*: The role of subtelomeric chromatin." Fly (Austin): 0.

Mason, J. M. and A. Villasante (2014). "Subtelomeres in *Drosophila* and Other Diptera." In Subtelomeres E. J. Louis and M. M. Becker (eds.), Springer Berlin Heidelberg: 211-225.

Muerdter, F., I. Olovnikov, A. Molaro, N. V. Rozhkov, B. Czech, A. Gordon, G. J. Hannon and A. A. Aravin (2012). "Production of artificial piRNAs in flies and mice." RNA 18(1): 42-52.

O'Kane, C. J. and W. J. Gehring (1987). "Detection in situ of genomic regulatory elements in *Drosophila*." Proc Natl Acad Sci U S A 84(24): 9123-9127.

Quesneville, H., C. M. Bergman, O. Andrieu, D. Autard, D. Nouaud, M. Ashburner and D. Anxolabehere (2005). r4 TE data.

Roche, S. E. and D. C. Rio (1998). "Trans-silencing by P elements inserted in subtelomeric heterochromatin involves the *Drosophila* Polycomb group gene, Enhancer of zeste." Genetics 149(4): 1839-1855.

Ronsseray, S., T. Josse, A. Boivin and D. Anxolabehere (2003). "Telomeric transgenes and trans-silencing in *Drosophila*." Genetica 117(2-3): 327-335.

Ronsseray, S., M. Lehmann and D. Anxolabehere (1989). "Copy number and distribution of P and I mobile elements in *Drosophila melanogaster* populations." Chromosoma 98(3): 207-214.

Ronsseray, S., M. Lehmann and D. Anxolabehere (1991). "The maternally inherited regulation of P elements in *Drosophila melanogaster* can be elicited by two P copies at cytological site 1A on the X chromosome." Genetics 129(2): 501-512.

Silva-Sousa, R. and E. Casacuberta (2013). "The JIL-1 kinase affects telomere expression in the different telomere domains of *Drosophila*." PLoS One 8(11): e81543.

Stanley, C. E., Jr. and R. J. Kulathinal (2016). "Genomic signatures of domestication on neurogenetic genes in *Drosophila melanogaster*." BMC Evol Biol 16(1): 6.

Todeschini, A. L., L. Teyssset, V. Delmarre and S. Ronsseray (2010). "The epigenetic trans-silencing effect in *Drosophila* involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1." PLoS One 5(6): e11032.

Walter, M. F., C. Jang, B. Kasravi, J. Donath, B. M. Mechler, J. M. Mason and H. Biessmann (1995). "DNA organization and polymorphism of a wild-type *Drosophila* telomere region." Chromosoma 104(4): 229-241.

Yin, H. and H. Lin (2007). "An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*." Nature 450(7167): 304-308.

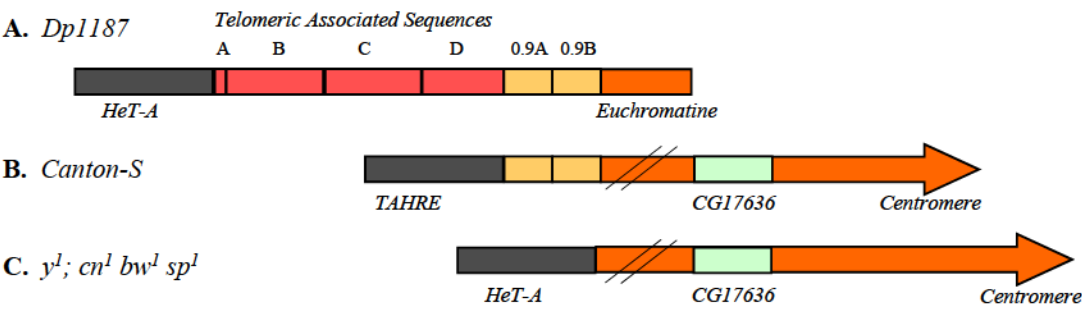


Figure 1

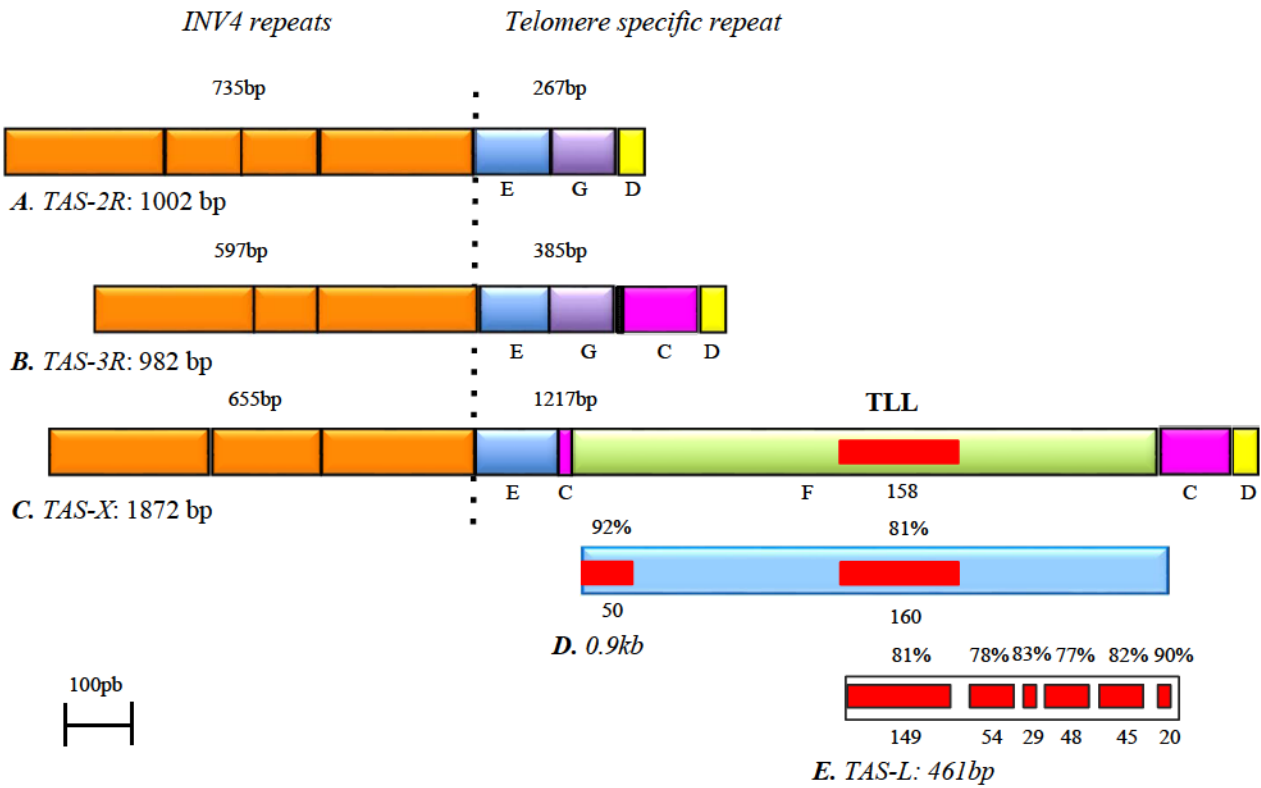


Figure 2

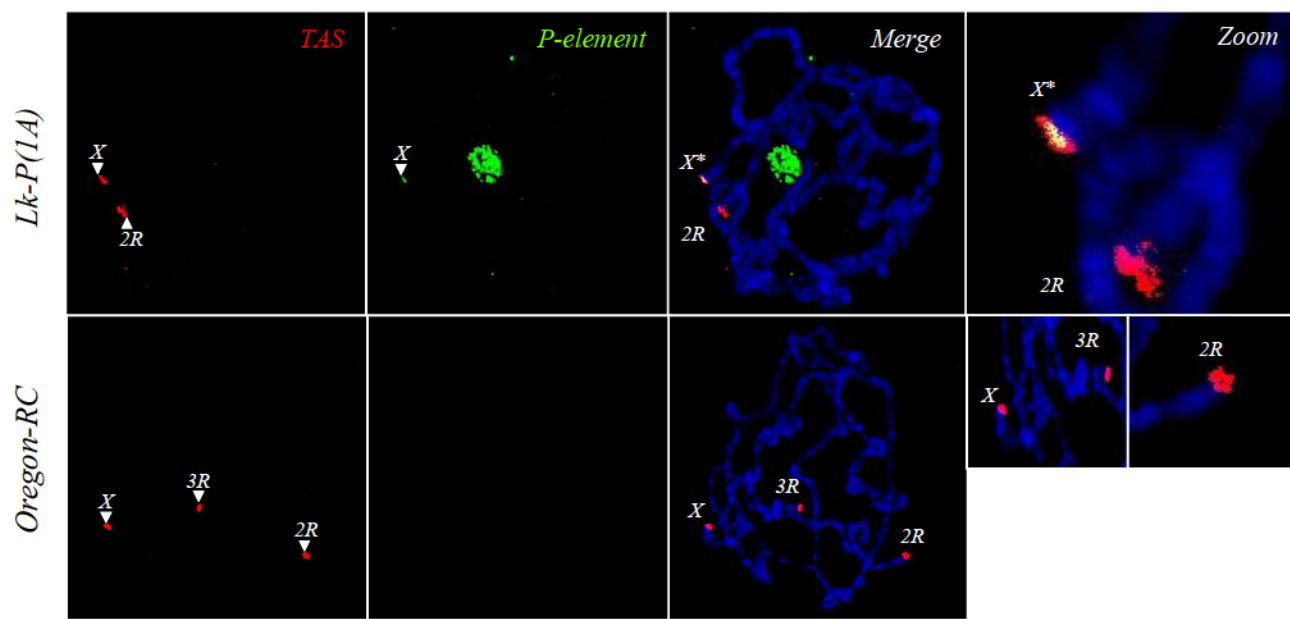


Figure 3

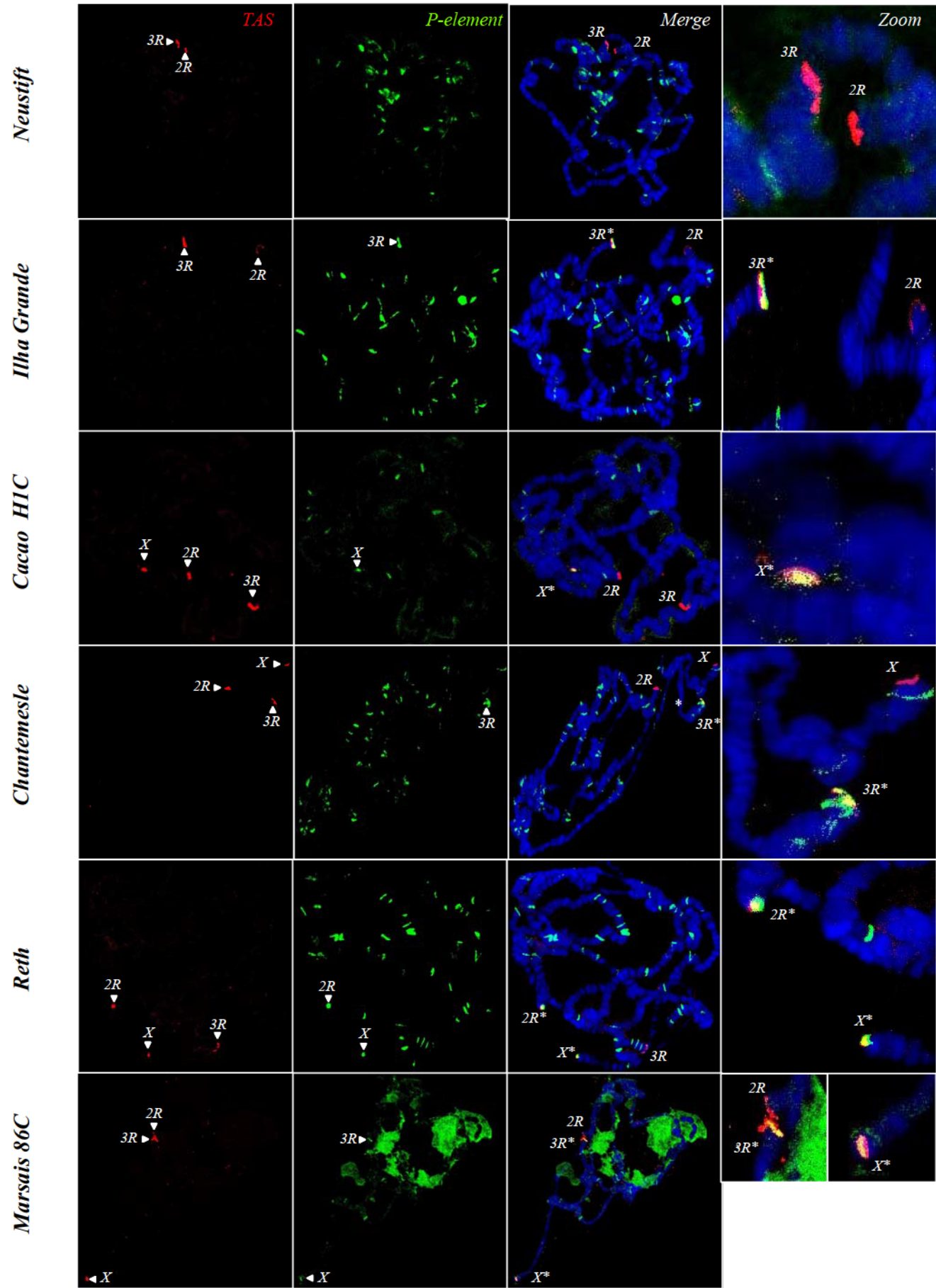


Figure 4

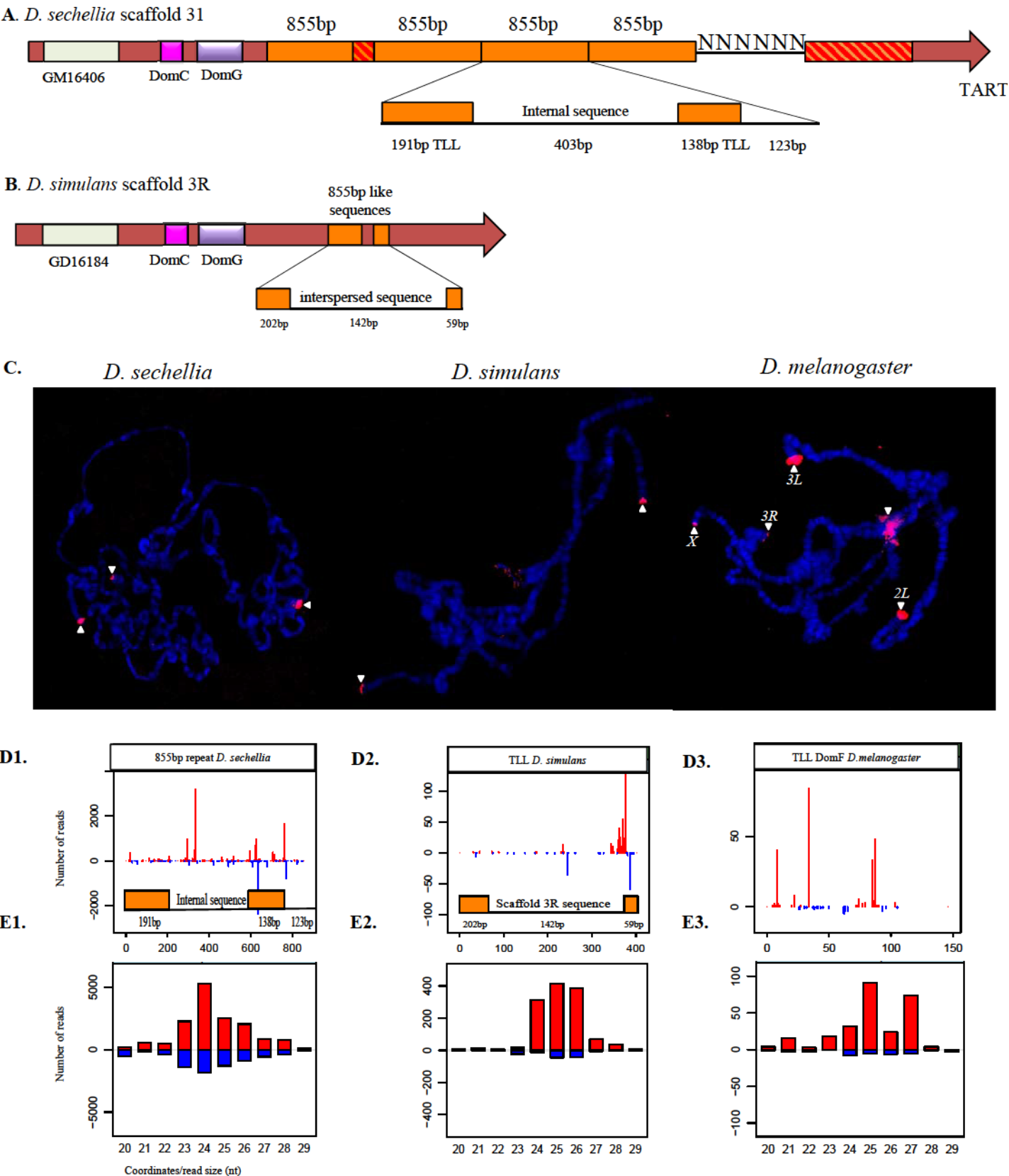


Figure 5

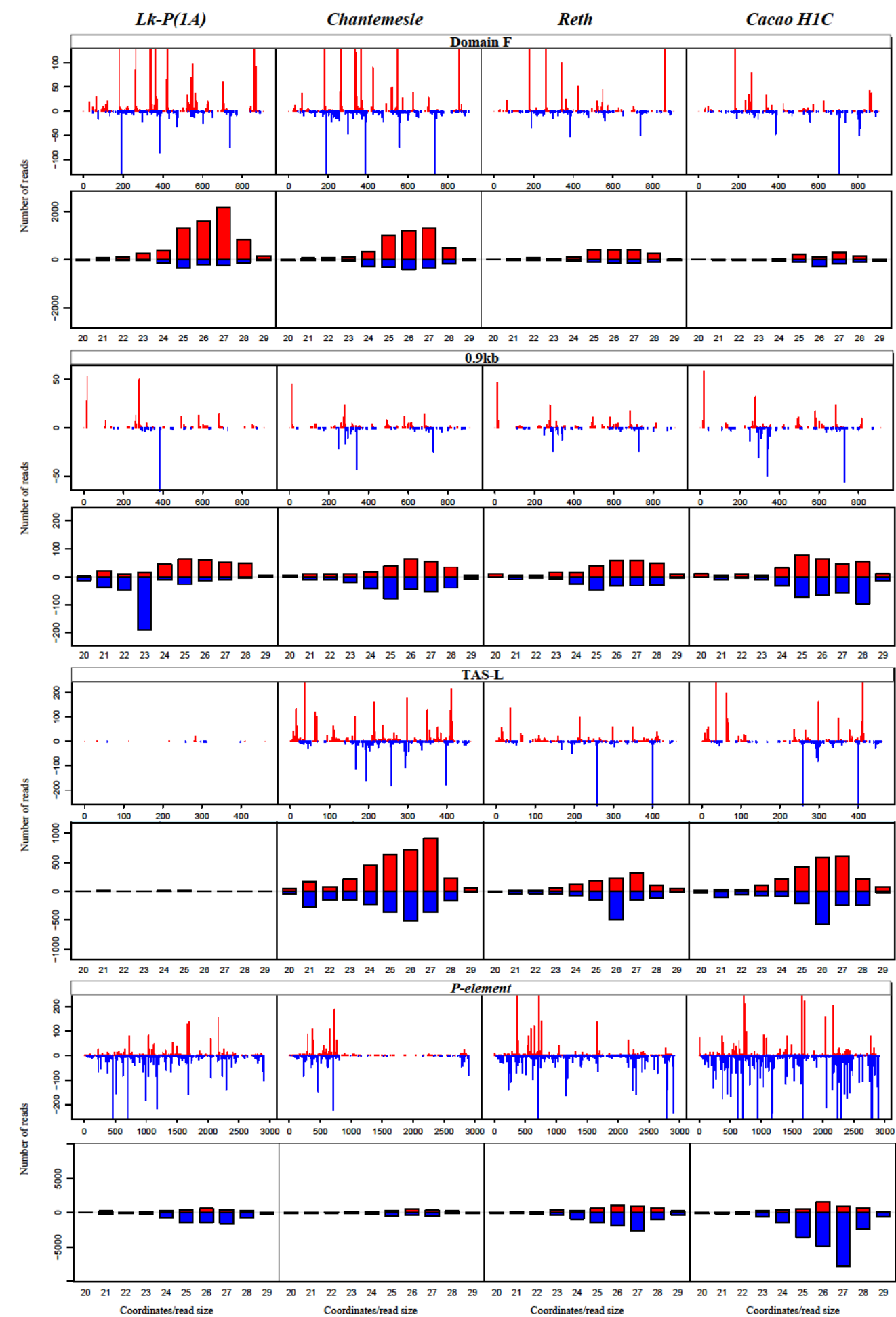
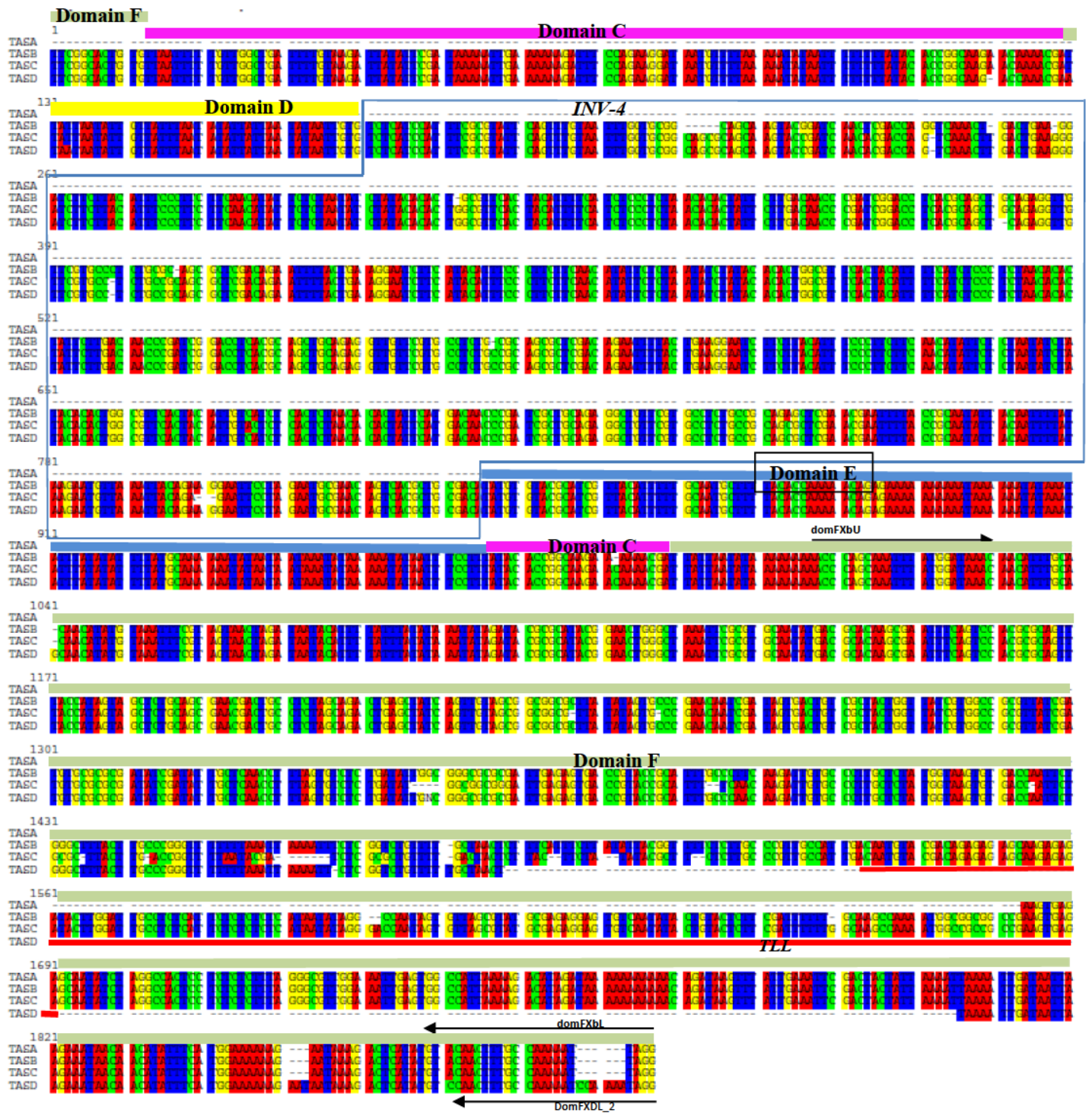
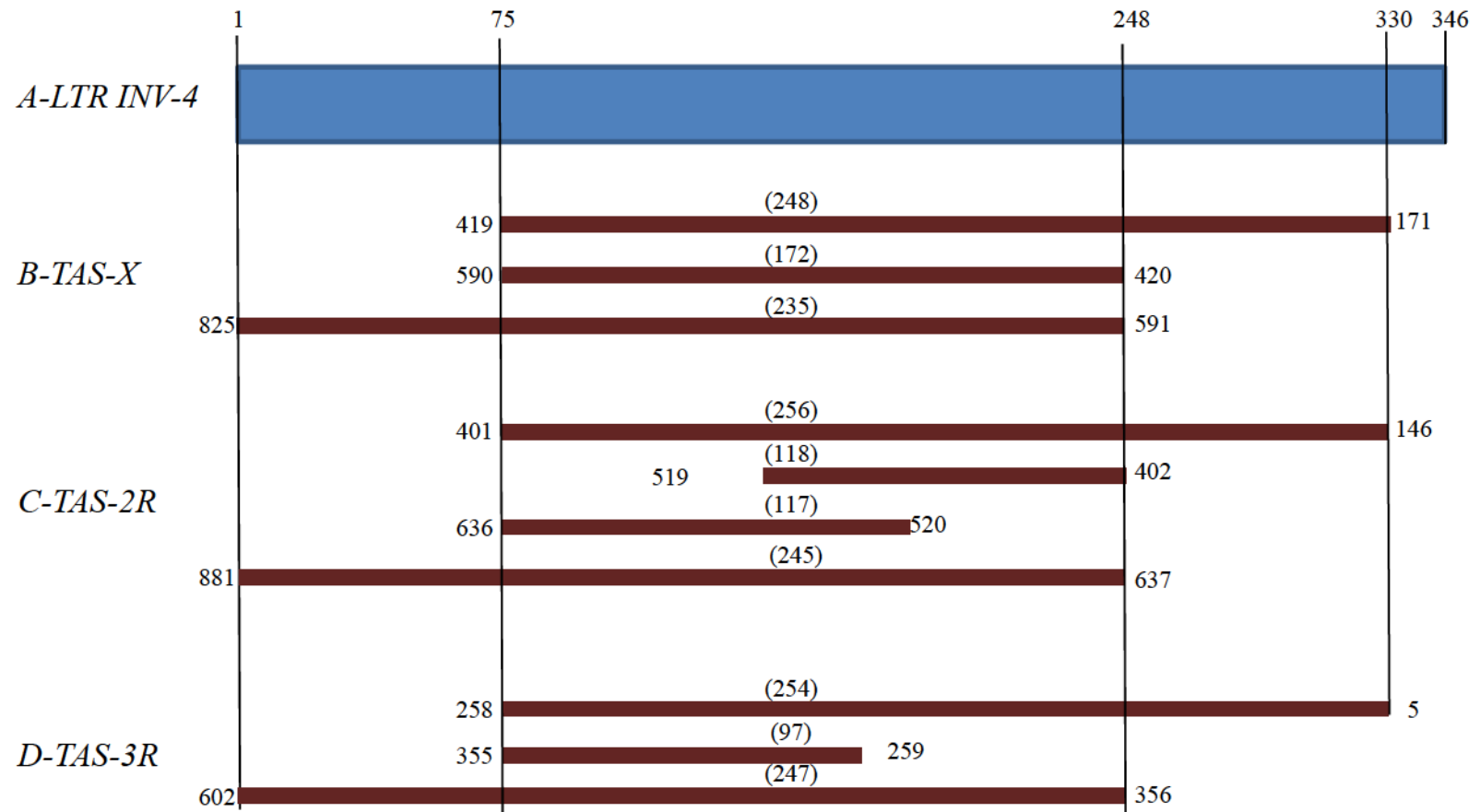
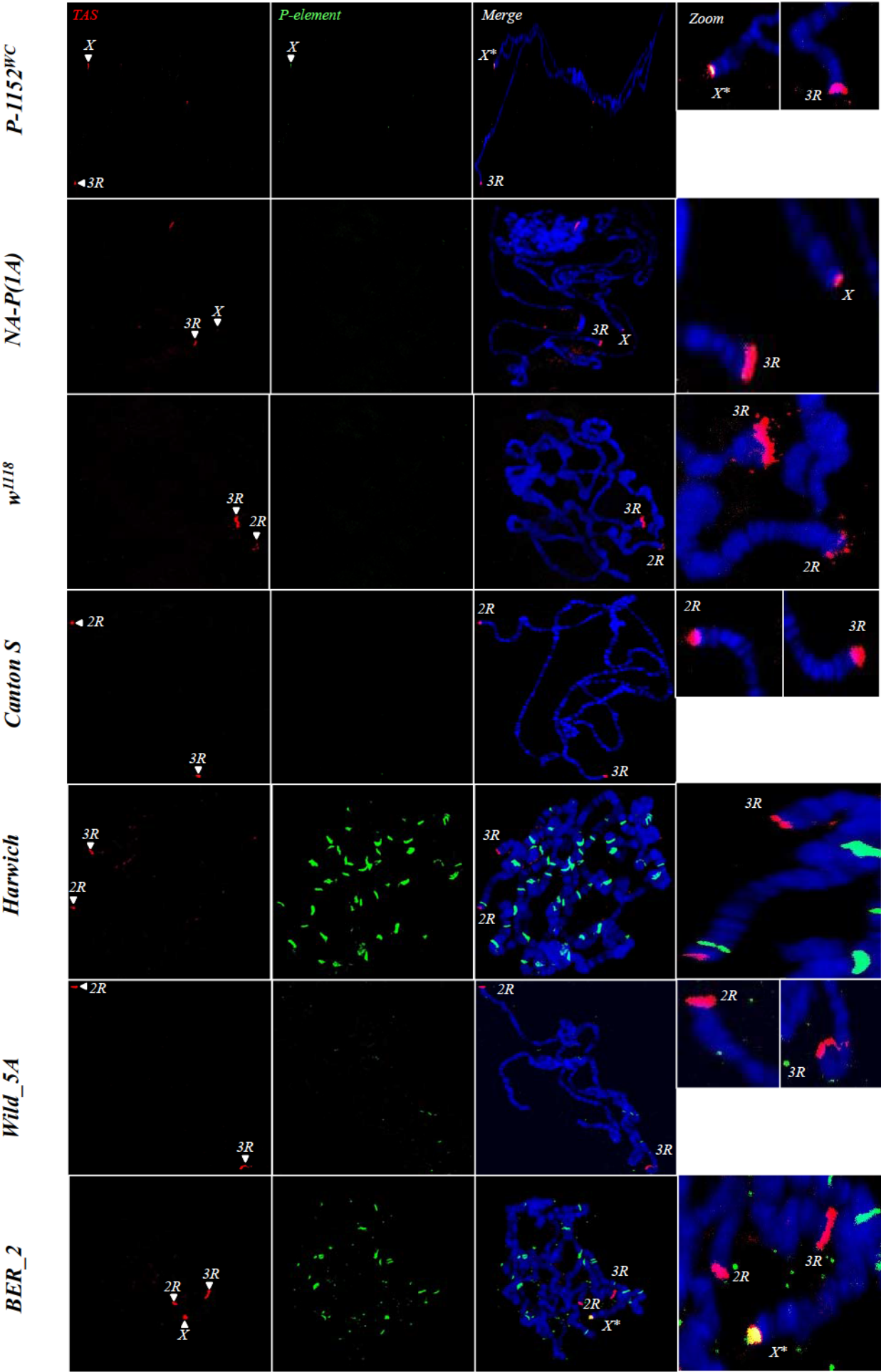


Figure 6

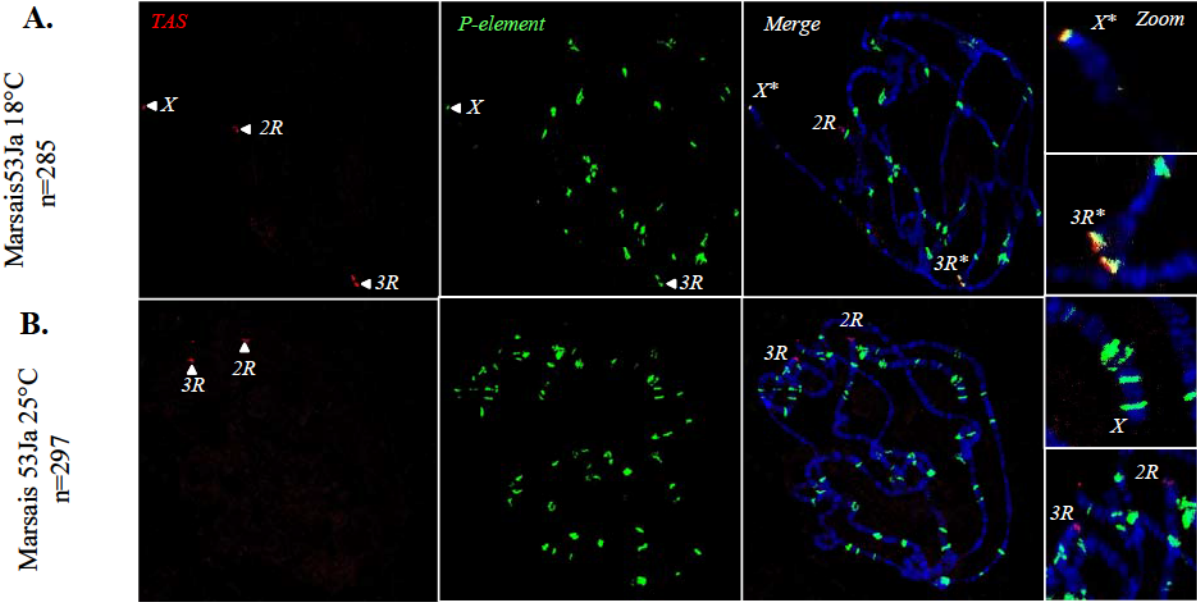


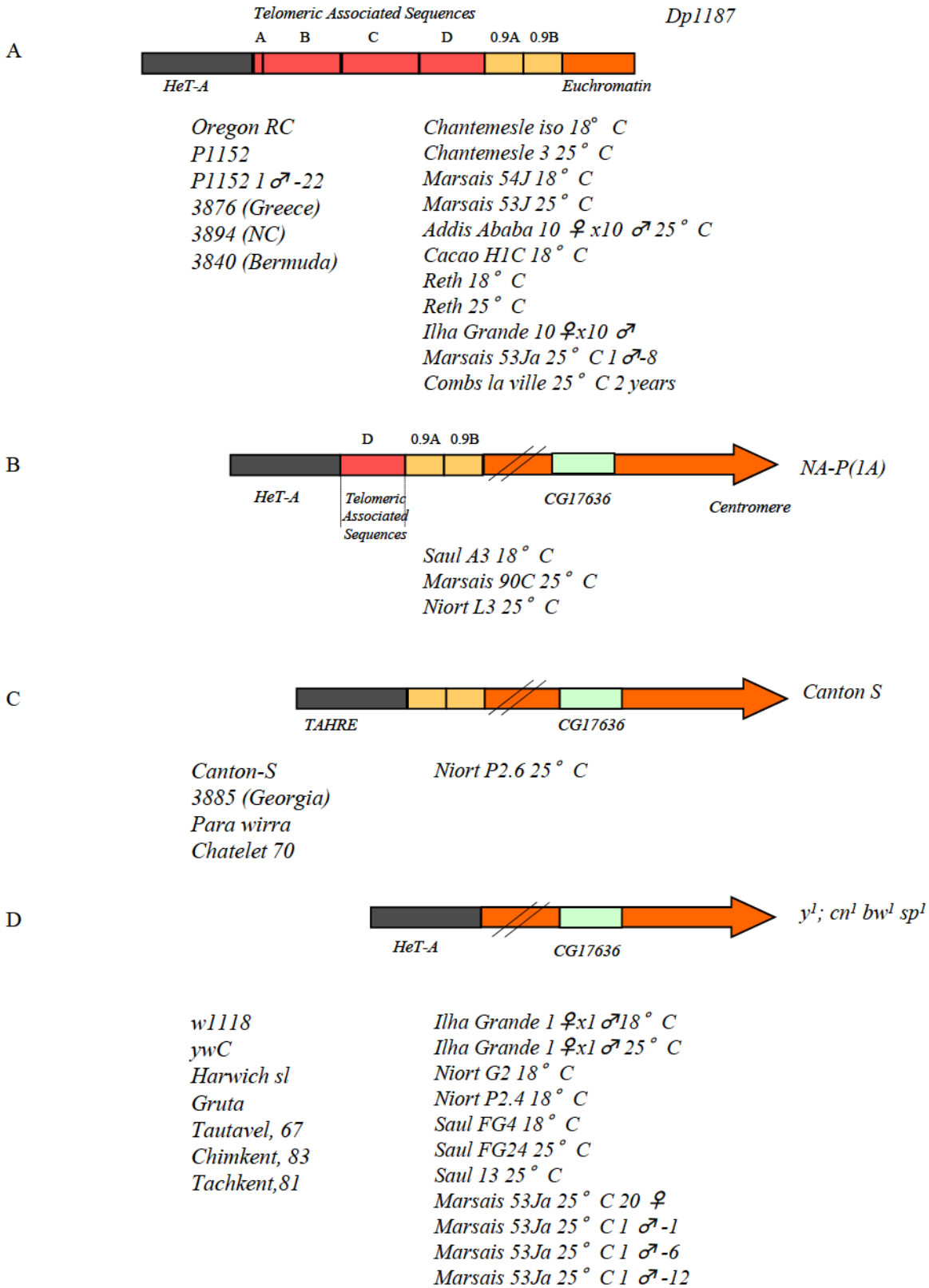
Supplementary figure 1

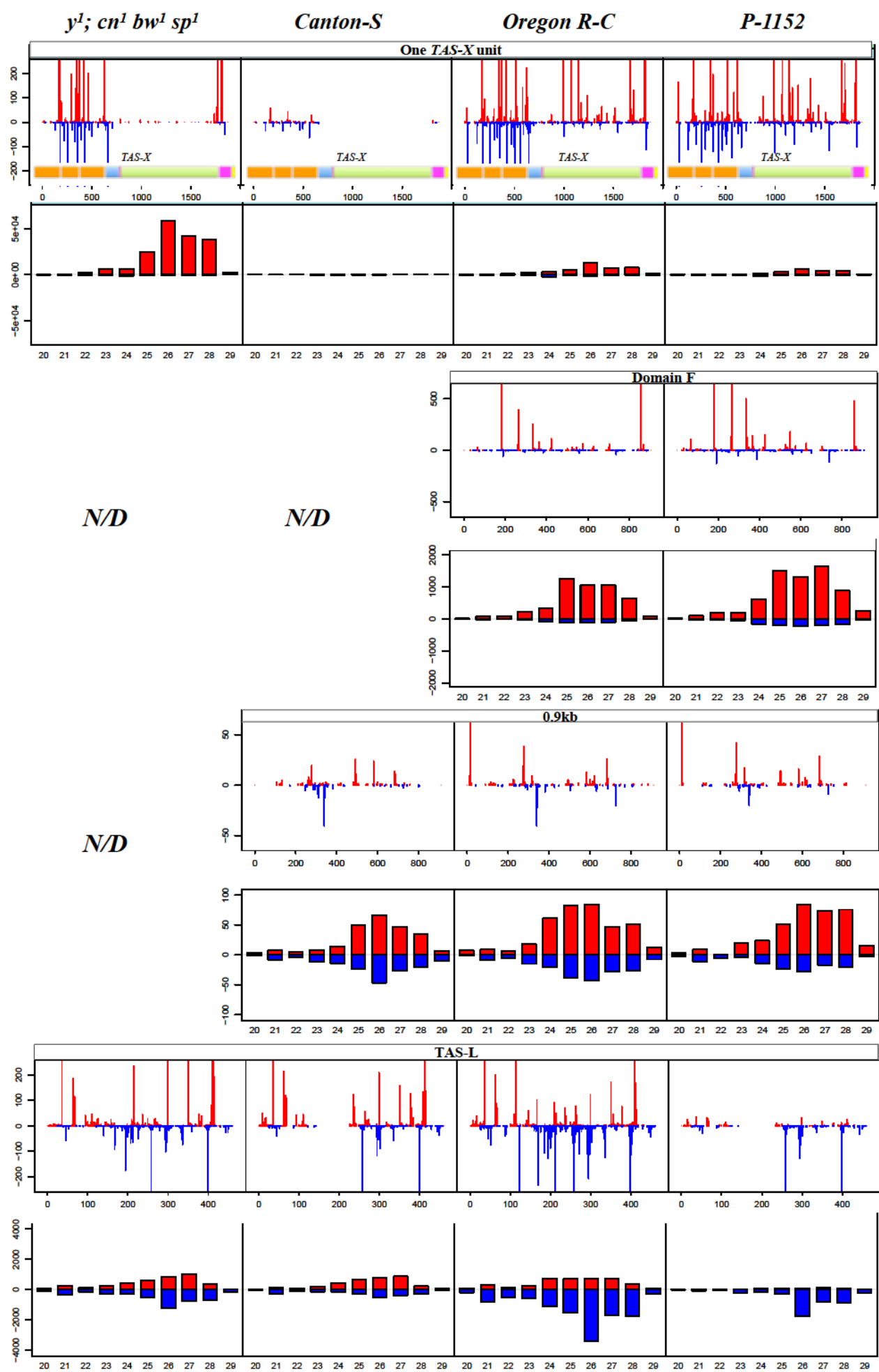




Supplementary figure 3

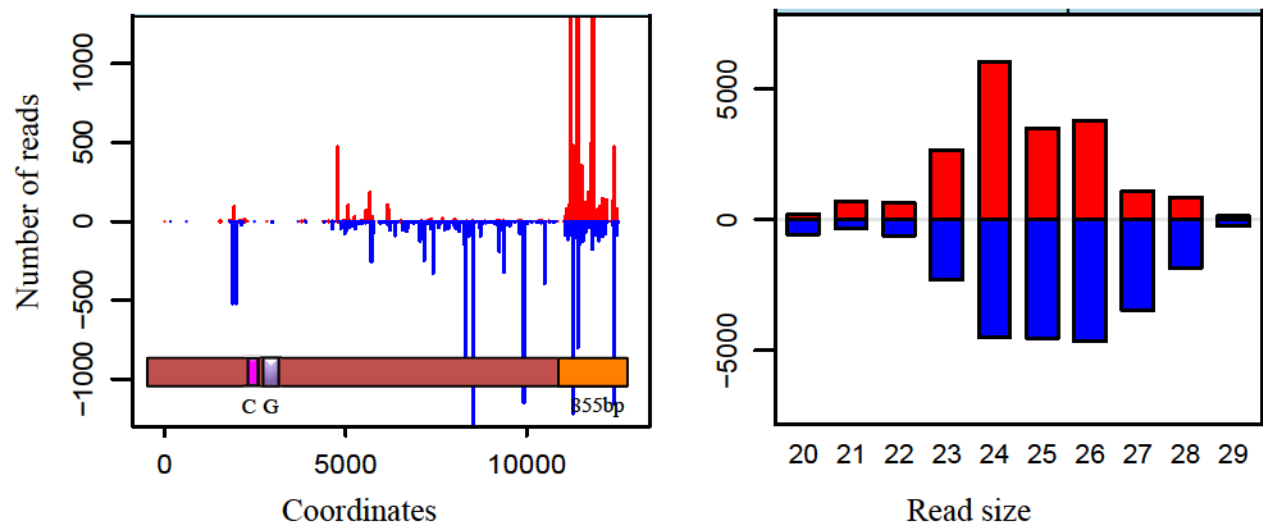






Supplementary figure 6

D. sechellia



D. simulans

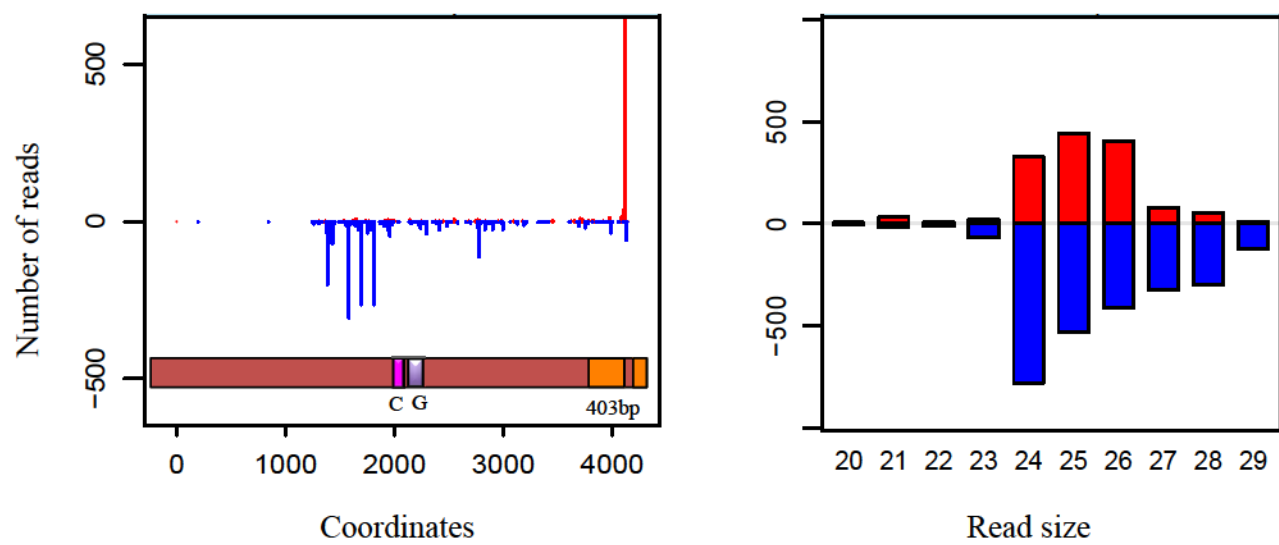


Table 1: Distribution of TAS-R domains in various laboratory strains

	Year of capture	<i>TAS-X</i> (1A)	<i>TAS-2R</i> (60F)	<i>TAS-3R</i> (100F)	PCR dom F	<i>P</i> element
<i>Lk-P(1A)</i> ^{a,c}	NA	+	+	-	+	<i>TAS-X</i>
<i>NA-P(1A)</i> ^{a,c}	NA	+	-	+	-	<i>TAS-X</i>
<i>P-1152</i> ^{wc a,c}	NA	+	-	+	+	<i>TAS-X</i>
<i>w</i> ^{1118 a}	NA	-	+	+	-	-
<i>yw</i> ^{c a}	NA	-	+	+	-	-
<i>Canton-S</i> ^b	1935	-	+	+	-	-
<i>Oregon R-C</i> ^b	1925	+	+	+	+	-
<i>y^l; cn^l bw^l sp^l</i> ^b	NA	-	+	+	-	-
Gruta ^a , France	1950-1955	-	+	-	-	
Tachkent ^a , Uzbekistan	1981	-	-	+	-	+
Tautavel ^a , France	1967	-	+	-	-	+
Para Wirra ^a , Australia	1972	-	+	+	-	-
Chatelet ^a , France	1970	-	+	+	-	+
Harwich ^a	1967	-	+	+	-	arms
Chimkent ^a , Kazakhstan	1983	NT	NT	NT	-	+
Wild_5A ^b , GA	1966	-	+	+	-	arms
VAG_2 ^b , Greece	1965	NT	NT	NT	+	+
Wild_11C ^b , NC	1966	NT	NT	NT	+	-
BER_2 ^b , Bermuda	1954	+	+	+	+	<i>TAS-X</i>

^a: Our laboratory stocks

^b: Bloomington stock center

^c These strains were used as control for TAS and *P* sequences

P element sequences were determined using *in situ* hybridization for chromosomal location (on arms or on subtelomeric position) or PCR amplification (+/-)

NA: Not Applicable because these strains are derived from other strains.

NT: Not Tested

Supplementary Table 1: Sequence of Primers used in PCR

Name	Oligo sequences	Used for
DomFXbU	AATCTAGACCCAGCAAATTTATGGATAAAC	F domain: <i>TAS-B</i> specific
DomFXbL	ATTCTAGACCTAATTTTTGGCAAAGTTGTAC	
P3-31	TGATGAAATAACATAAGGTGGTCCCGTCG	Full length <i>P</i> element
DomFXDL_2	CCTATTTTGGATTTTGGCAAAGTTG	<i>TAS-D</i> specific with DomFXbU
0.9L	TGGGAGGAATAATGTGGAGCA	0.9 kb repeat specific
0.9R	CGCCCTGTTTTCTCCTATTGG	
<i>SerepU</i>	GAGGAGTGGCCTAGTTATTGC	<i>Drosophila sechellia</i> repeats
<i>SerepL</i>	CCATATTGCGACCCTAAGTAGT	

Supplementary Table 2: Annotation of small RNA libraries

Small RNAs were prepared from ovaries of females of the indicated strains. Values for the different categories of sequences matching the *D. melanogaster* genome 5.49 in each library are indicated. For comparisons, libraries were normalized (normalization factor) relatively to the number of sequence reads aligning to the *D. melanogaster* genome but not to tRNAs or miscRNAs (effective depth). Library #1 has the lowest effective depth and was taken as reference to normalize the other libraries. The normalization factor for *D. sechellia* and *D. simulans* are not shown since all the data are not available for libraries annotations.

Library	1	2	3	4	5	6	7	8
Strain	<i>LK-P(1A)</i>	<i>Canton^y</i>	<i>P-1152</i>	<i>y1; cn1 bw1 sp1</i>	<i>Oregon R-C</i>	<i>Chantemesle</i>	<i>Reth</i>	<i>Cacao H1C</i>
<i>D.mel</i> matched (R5.49)	21672810	10153341	10011295	45090091	11809532	10673891	9696015	9368805
<i>D.mel</i> _all-miscRNA	15524554	924278	1416324	2999936	903167	1947337	1458180	1462296
<i>D.mel</i> _all-ncRNA	29415	701997	64608	3756508	934875	687906	636664	673753
<i>D.mel</i> _all-tRNA	67375	107897	557997	730316	327534	106519	156815	416052
<i>D.mel</i> _all-transposon	3345426	4423806	4067114	23765534	5310749	4386203	4214248	3509012
<i>D.mel</i> _miR_r20	848754	1638666	1235457	4654111	1674572	1341836	1244774	1615109
<i>D.mel</i> _all-intergenic	1086741	1656792	1404861	6377523	1752717	1536649	1373326	1189880
<i>D.mel</i> _all-intron	108622	222188	206044	757035	242278	197332	179893	14302
Remaining Unmatched	397188	477717	477418	2049128	66364	470109	432115	359683
Effective depth	6080881	9121166	8036974	41359839	10578831	8620035	8081020	7490457
Normalization Factor	1	0,67	0,76	0,15	0,57	0,71	0,75	0,81

Supplementary Table 3. Distribution of TAS-R in recently collected *Drosophila* strains

origin	Year	Comments	In situ hybridization			TAS and P colocalization	Few generations after trapping from wild ~G5	18°C		25°C	
			TAS-X (1A)	TAS-2R (60F)	TAS-3R (100F)			1 year (~G9)	2 years (~G20- 24)	1 year (~G20- G23)	2 years (~G41- 44)
Europe											
France											
Ile de France											
Paris (75)	2013	isofemale	+	+	-	NT	+		+	+	+
Combs-la-Ville (77)	2013	isofemale	+	+	+	NT	+		+	+	+
Viry-Châtillon (91) b	2013	isofemale						+	+		
Viry-Châtillon (91) c	2013	isofemale	+	+	+	NT		+	+		
Chantemesle (95) - iso	2013	isofemale	+	+	+	TAS-3R	+		+		+
Chantemesle (95) - 1	2013	offspring					+		+		
Chantemesle (95) - 2	2013	offspring					+		+	+	+
Chantemesle (95) - 3	2013	offspring						+	+	+	+
Aquitaine, Poitou, Limousin											
Marsais (17) -47J	2013	isofemale					+	+	+	+	+
Marsais (17) -53Ja	2013	isofemale	+	+	+	TAS-X; TAS-3R		+	+	+	-
Marsais (17) -54J	2013	isofemale						+	+	+	+
Marsais (17) -57C	2013	isofemale						+	+	+	+
Marsais (17) -58C	2013	isofemale						+	+	+	+
Marsais (17) -66C	2013	isofemale						+	+		+
Marsais (17) -73J	2013	isofemale						+	+	+	+
Marsais (17) -80J	2013	isofemale	+	+	-	NT		+		+	+
Marsais (17) -81J	2013	isofemale								+	+
Marsais (17) -82J	2013	isofemale						+		+	+
Marsais (17) -86C	2013	isofemale	+	+	+	TAS-X; TAS-3R		+	+	+	+
Marsais (17) -87C	2013	isofemale	+	+	-	NT		+	+	+	+
Marsais (17) -90C	2013	isofemale						+	+	+	-
Niort (79) - A1	2013	offspring	+	-	+	NT	+		+	+	+
Niort (79) - A2	2013	offspring					+		+	+	+
Niort (79) - A3	2013	offspring					+		+	+	+
Niort (79) - A4	2013	offspring							+	+	+
Niort (79) - G1	2013	isofemale					+	+		+	+
Niort (79) - G2	2013	isofemale						+	-	+	+
Niort (79) - G3	2013	isofemale	+	+	-	NT	+	+	+		
Niort (79) - G4	2013	isofemale						+		+	+
Niort (79) - P1.1	2013	isofemale						+	+	+	+
Niort (79) - P1.2	2013	isofemale						+	+	+	+
Niort (79) - P1.3	2013	isofemale	+	-	+	NT	+	+	+	+	+
Niort (79) - P1.4	2013	isofemale						+	+	+	+
Niort (79) - P1.5	2013	isofemale						+	+	+	+
Niort (79) - P1.6	2013	isofemale						+	+	+	+
Niort (79) - P2.1	2013	isofemale	+	-	+	NT	+	+	+	+	+
Niort (79) - P2.2	2013	isofemale								+	
Niort (79) - P2.3	2013	isofemale									+
Niort (79) - P2.4	2013	isofemale						+	-	+	+
Niort (79) - P2.5	2013	isofemale						+	+	+	+
Niort (79) - P2.6	2013	isofemale						+		+	-
Niort (79) - L1	2013	isofemale						+	+	+	+
Niort (79) - L2	2013	isofemale							+	+	+
Niort (79) - L3	2013	isofemale								+	-
Reth (85)	2013	isofemale	+	+	+	TAS-X; TAS-2R	+		+	+	+
Mauzé Thouarsais (79) - 1	2013	offspring					+		+	+	+
Mauzé Thouarsais (79) - 2	2013	offspring	+	-	+	NT	+		+		
Mauzé Thouarsais (79) - 3	2013	offspring	+	-	+	NT	+		+		+
Mauzé Thouarsais (79) - 4	2013	offspring							+		
Mauzé Thouarsais (79) - 5	2013	offspring							+	+	+
Mauzé Thouarsais (79) - 6	2013	offspring							+	+	+
Mauzé Thouarsais (79) - 7	2013	offspring							+	+	+
Sanguinet (40) - 1.1	2014	isofemale					+	+		+	
Sanguinet (40) - 1.2	2014	offspring					+	+		+	
Sanguinet (40) - 1.3	2014	offspring					+	+		+	
Sanguinet (40) - 1.4	2014	offspring					+	+		+	
Sanguinet (40) - 1.5	2014	offspring	+	+	+	none	+	+		+	
Sanguinet (40) - 1.6	2014	offspring					+	+		+	
Sanguinet (40) - 3	2014	isofemale					+	+		+	
Val de Loire											
Rouans (44) - 1	2013	isofemale	+	+	+	NT	+		+	+	+
Rouans (44) - 2	2013	isofemale					+		+	+	+
Auvergne Rhône Alpes											
Gotheron, Valence (26) -ISO3	2012	isofemale	+	+	-	NT		+	+	+	+
Austria											
Mannsdorf an der Donau	2014	isofemale					+				
Neustift am Walde (Vienna)	2014	isofemale	-	+	+	none	+	+		+ G15	
Africa											
Ethiopia											
Addis Ababa	1F x 1M	2014	isofemale				+	+		+ G9	
	10F x 10M	2014	+	+	+	NT	+	+		+ G9	
Senegal											
Kafountine	2015	offspring	+	+	-	TAS-2R	+			+ G6	
South America											

French Guiana (973)									
Saul, FG 4	dec, 2013	isofemale						+	+
Saul, FG24	dec, 2013	isofemale					+a	+ G17	+
Saul, FG31	dec, 2013	isofemale					+a	+ G17	+
Saul, FG34	dec, 2013	isofemale					+a	+ G20	+
Saul, Cacao H2A1	July, 2014	isofemale					+a	+* G18	+
Saul, Cacao H2A2	July, 2014	isofemale	+	+	-	NT	+a	+ G18	+
Saul 1A	July, 2014	isofemale					+a		+
Saul A2	July, 2014	isofemale					+a		+
Saul A3	July, 2014	isofemale					+a	- G18	+
Saul, Cacao H1C	July, 2014	isofemale	+	+	+	TAS-X	+a	+ G18	+
Saul A1	July, 2014	isofemale					+	+ G18	+
Saul B	July, 2014	isofemale					+	+ G18	+
Saul C	July, 2014	isofemale					+	+ G18	+
Saul 1 2014	Sept, 2014	isofemale					+	+	+
Saul 2 2014	Sept, 2014	isofemale					+		
Saul 3 2014	Sept, 2014	isofemale					+		
Saul 4 2014	Sept, 2014	isofemale					+		+
Saul 6 2014	Sept, 2014	isofemale					+		
Saul 8 2014	Sept, 2014	isofemale					+		
Saul 9 2014	Sept, 2014	isofemale					+		
Saul 10 2014	Sept, 2014	isofemale					+		
Saul 11 2014	Sept, 2014	isofemale					+		
Saul 12 2014	Sept, 2014	isofemale					+		+
Saul 13 2014	Sept, 2014	isofemale					+	+*	-
Saul 14 2014	Sept, 2014	isofemale					+	+	
Saul 16 2014	Sept, 2014	isofemale					+	+	+
Brazil									
Ilha Grande, Rio	1♀ x 1♂	2014	isofemale	-	+	+	TAS-3R	-*	-
Janeiro	10♀ x 10♂	2014						-*	+

"Comments" column

Isofemale When a unique inseminated female was collected from wild and used to established a stock

Offspring When strains were established from 3-5 males and females emerged from a trap containing fruits,

"*in situ* hybridization" column: Hybridization on polytene chromosomes were performed using the 1.8-kb TAS probes, some of the results are presented in **Supplementary Fig. 4** and **Fig. 5**

+, positive hybridization on TAS. -, no hybridization on TAS

"TAS and P colocalization" column: Hybridization on polytene chromosomes were performed on some strains using a 2.9kb P probe, some of the results are presented in **Supplementary Fig. 4** and **Fig. 5**

In all strains, P hybridization signals were observed on chromosomal arms. In some strains hybridization signals were observed on telomeric loci.

When colocalization between TAS locus and P element is observed, the locus of the colocalization is specified

"none": when P elements are present on chromosomal arms but no colocalization with TAS locus are observed

PCR amplifications were performed at least 2 times. +, positive amplification. -, negative amplification, +*: faint band observed

a: when strains were tested after few generations after their capture (between 3 to 5)

: when strain was lost due to unknown decreasing fertility

"*": the strains collected in Addis Ababa (Ethiopia) and Ilha Grande (Brazil) were first tested 3 months after their capture. In case, some polymorphisms were already existing, we established a strain by crossing one virgin female with one male (1♀x1♂), and another strain by crossing 10 males with 10 virgin females (10♀x10♂)

NT: not tested

Red letters were used to highlight strains that lost TAS-X locus

Supplementary Table 4. Loss of *TAS-X* in 2 years period of laboratory condition

20 males of the Marsais 53Ja strain (G46) were individually PCR amplified using primers specific for *TAS-X*

	PCR dom F
20 ♀ G46 25°C	-
♂1	-
♂2	-
♂3	-
♂4	-
♂5	-
♂6	-
♂7	-
♂8	+
♂9	-
♂10	-
♂11	-
♂12	-
♂13	-
♂14	-
♂15	-
♂16	-
♂17	-
♂18	+
♂19	-
♂20	-
1 ♂ w1118	-
1 ♂ P1152	++

PCR on one male of the w1118 and P1152 strains were performed as negative and positive control, respectively

Conclusions

L'étude de la structure du locus TAS a permis de mettre en évidence que les trois TAS-R présents aux chromosomes *X*, *2R* et *3R* sont plus proches que ce qu'il avait pu être décrit précédemment dans la littérature. Une séquence de 158bp, que nous avons appelée *TLL*, commune aux TAS-R et aux TAS-L, capable de produire des piRNA, est retrouvée chez d'autres espèces du groupe *melanogaster* nous permettant d'émettre l'hypothèse que cette séquence peut être à l'origine des TAS. Par ailleurs, l'étude de la présence de TAS dans les souches de laboratoires et dans les souches de la nature montre premièrement qu'il n'existe aucune souche de drosophiles dépourvues de séquences TAS aux trois télomères, deuxièmement, deux locus TAS ont toujours pu être observés, troisièmement que les populations naturelles possèdent toutes un *TAS-X* et enfin ce *TAS-X* peut être perdu au cours des générations quand ces souches sont maintenues dans les conditions de laboratoires. Cette analyse met donc en évidence un rôle particulier du *TAS-X* dans le génome des drosophiles de la nature par rapport aux souches conservées dans les laboratoires.

2. Analyse de mutants pour la production de piRNA du locus TAS

Au laboratoire, nous possédons un certain nombre de données de séquençage issu de divers contexte génomiques. Parmi celles-ci, il existe une série de banque de petits ARN issus de souches mutantes pour *Aubergine*, *Rhino*, *Cutoff* et *Zucchini* qui sont des gènes impliqués dans la voie de biogenèse des piRNA. Ces banques ont été générées lors de l'étude du phénomène de la paramutation (Hermant *et al.* 2015) que je vous présenterai dans la partie suivante. Nous possédons également une banque de petits ARN issus d'un mutant demi-dose (la souche homozygote n'étant pas viable) pour le gène *Su(var)205* codant la protéine HP1, composante de l'hétérochromatine. En effet, la répression des transgènes *P-lacZ* télomérique est affectée en conditions mutantes pour ce gène (Josse et al, 2007). Cependant la production des piRNA n'avait pas été testée à l'époque et notamment ceux issus du locus TAS où ce transgène est inséré. Ces banques ont donc été utilisées afin d'analyser l'impact de ces différents mutants sur la production des piRNA (23-29nt) issus d'une part des TAS-R et d'autre part des TAS-L. Le but de l'analyse décrite ici ne sera pas de détailler les effets précis de chacun de ces mutants mais d'analyser leur impact de façon globale.

A1. Les TAS-R

Les lignées de drosophiles utilisées pour établir les mutants pour les gènes *Aubergine* (*Aub*), *Cutoff* (*Cuff*), *Rhino* et *Zucchini* (*Zuc*) ne possèdent pas de *TAS-X* mais un *TAS-2R* ainsi qu'un *TAS-3R*. En conséquence, c'est un domaine commun entre ces deux TAS, le domaine G qui a été utilisé comme séquence de référence. Ce domaine reflètera donc la production des piRNA des TAS-R. Par ailleurs, la souche mutante pour *Su(var)205* possède quant à elle le *TAS-X*. Il s'agit de la souche *Lk-P(1A)* contenant deux éléments *P* insérés dans le *TAS-X* (voir article I). Dans ce cas, la séquence du domaine F, séquence uniquement présente dans le *TAS-X* a pu être utilisée (voir article I).

La production de piRNA du domaine G est fortement impactée en conditions mutantes pour *Rhino* et *Cuff* (Figure 28E et 28H). Ces deux gènes codent pour les protéines Rhino et Cutoff faisant partis du complexe Rhino/Deadlock/Cutoff (RDC) définissant la transcription non canonique des dual-strand clusters (Mohn *et al.* 2014). Ce complexe induit l'absence de terminaison de transcription ainsi l'inhibition de l'épissage. La transcription des TAS-R semble donc être soumise au complexe RDC. Le même effet est observé pour le mutant *zuc*

(Figure 28J et 28K). En effet, cette protéine prend en charge les transcrits primaires qui sortent du noyau et à la base de la production des piRNA primaires. Par ailleurs, dans cette souche mutante, des piRNA du domaine F ont pu être détectés. Ceci indique donc la présence d'un *TAS-X* dans cette souche, absent de la souche sauvage, qui a pu être apporté par les chromosomes des souches possédant les allèles mutants pour *zuc*. La production de piRNA est aussi affectée pour le domaine F en conditions mutantes pour *zuc*. Pour le mutant *aub*, il y a une diminution de production de piRNA sans une perte totale (Figure 28B). La protéine Aubergine étant impliquée dans la biogenèse des piRNA secondaires, les piRNA observés peuvent être des piRNA primaires. Enfin, en ce qui concerne le mutant *Su(var)205*, une demi-dose suffit pour diminuer considérablement la quantité de piRNA aussi bien pour le domaine F que le domaine G (Figure 28M et 28N).

A2. Les TAS-L

Seuls les mutants pour *Rhino* et *Cuff* induisent une perte totale des piRNA issus des TAS-L (Figure 28F et 28I). Ces TAS seraient donc aussi pris en charge par le complexe RDC et auraient une transcription non canonique comme les dual-strand clusters et les TAS-R. Cependant, pour les mutants pour *Aub*, *Zuc* et *Su(var)205*, les effets sont beaucoup plus discutables. Pour *zuc*, seuls les piRNA en sens sont affectés (Figure 28L). Pour *Aub* et *Su(var)205* (Figure 28C et 28O), les effets sont inverses car la quantité de piRNA semble augmenter en conditions mutantes pour des raisons inconnues.

Il existe actuellement deux catégories de grandes clusters de piRNA, les uni-strand et les dual-strand cluster définies par leur présence respectivement en lignée somatique et en lignée germinale, leur sens de transcription ainsi que les voies biogenèse des piRNA qui leur sont associées. Ces descriptions ont été effectuées par l'analyse des clusters représentatifs des uni-strands (*Flamenco*) et dual-strand clusters (*42AB*). Il peut cependant exister des exceptions à ces règles comme c'est le cas du *cluster 2* (*20A*) qui est un uni-strand cluster mais présent en lignée germinale (Mohn *et al.* 2014). L'analyse des piRNA en conditions mutantes pour les TAS-L semble indiquer une biologie différente des piRNA par rapport aux TAS-R.

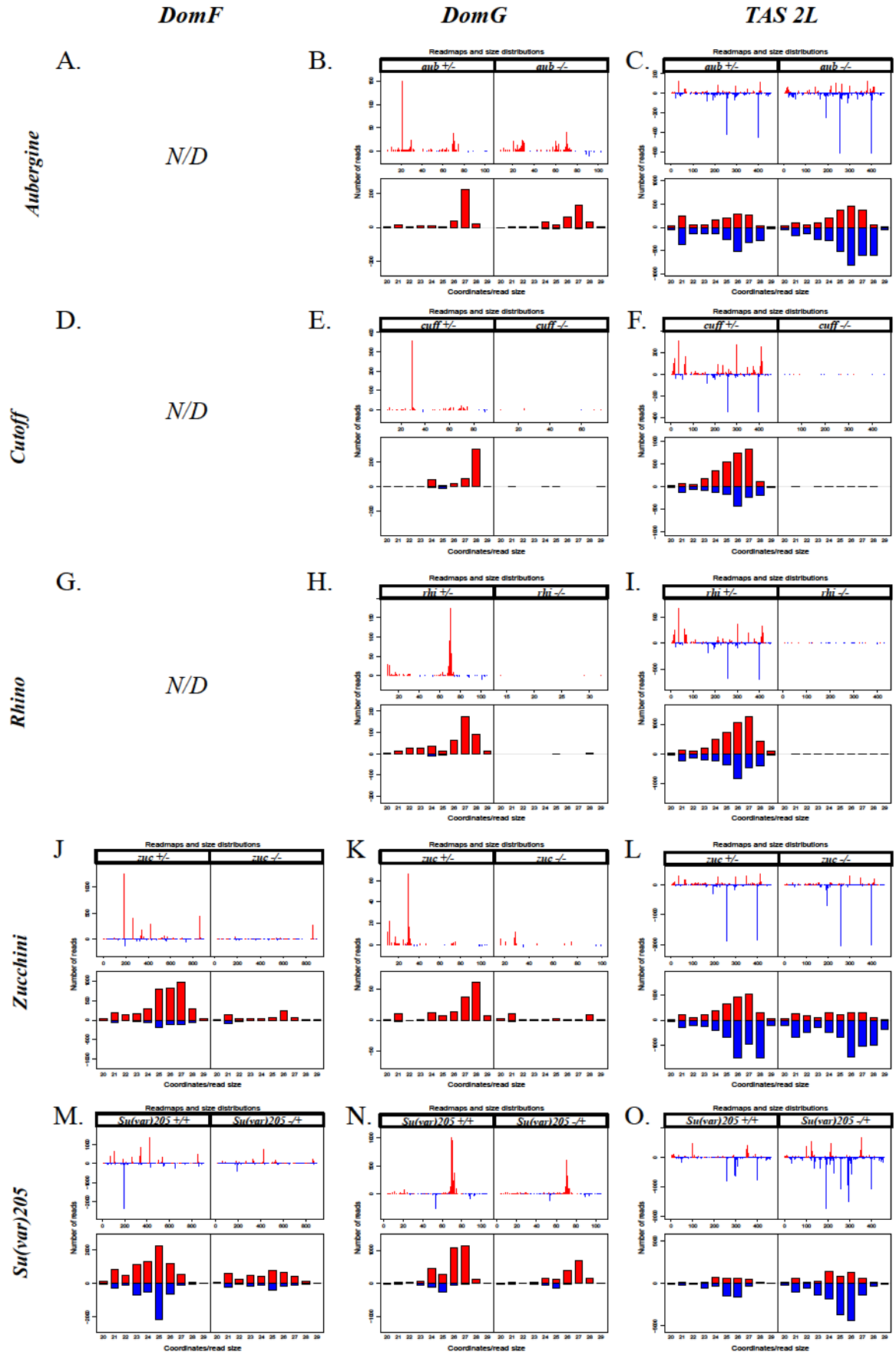


Figure 28. Distribution des petits ARN (20-29nt) sur le domaine F, le domaine G et le TAS-L pour les mutants *Aubergine*, *Cutoff*, *Rhino*, *Zucchini* et *Su(var)205*.

3. Article II

Présentation de l'article II

Au laboratoire un phénomène de conversion épigénétique appelé « paramutation » a été mis en évidence (de Vanssay et al, 2012). Elle met en jeu deux lignées *T-1* et *BX2* possédant un cluster de 7 transgènes *P-lacW* marqués par le gène responsable de la pigmentation de l'œil (white) insérés au site *50C* du chromosome 2 dans une région euchromatique. La lignée *T-1* produit des piRNA du cluster *P-lacW* alors que *BX2* n'en produit pas. Il a été montré que la lignée *T-1* peut transmettre ses propriétés épigénétiques (piRNA) à la lignée *BX2*. La souche *BX2* produit à son tour des piRNA du cluster *P-lacW*, la souche paramutée est appelée *BX2**. De plus, il a été montré que cette paramutation est stable à travers les générations et qu'une souche *BX2** peut à son tour transmettre ses propriétés épigénétiques à une souche *BX2* ne produisant pas de piRNA. Par ailleurs dans cette étude, il a été montré que les clusters *P-lacW* en plus de produire des piRNA pouvait aussi produire des siRNA.

Lors de mon arrivée au laboratoire, un projet concernant de l'analyse des propriétés fonctionnelles et moléculaires associées à cette paramutation était en cours. L'effet de mutants impliqués dans la voie de biogénèse des piRNA primaires (*Rhino*, *Cutoff* et *Zucchini*) et des piRNA secondaires (*Aubergine*) a été analysé sur la paramutation. L'effet d'un mutant pour Dicer-2, protéine impliquée dans la voie de biogénèse des siRNA a été aussi testée afin d'étudier l'implication des siRNA de *P-lacW* retrouvés dans la paramutation. Sont-ils nécessaires pour établir cette paramutation ou sont-ils un sous-produit issu de cette conversion épigénétique ? Enfin, l'effet de ces piRNA a aussi été testé dans les cellules somatiques car il a été observé un effet sur la pigmentation de l'œil drosophile lorsque les piRNA homologues au cluster *P-lacW* sont apportés maternellement.

Cette étude à laquelle j'ai aussi participé, a fait l'objet d'une publication en 2015 dans la revue *Genetics* (Hermant *et al.* 2015). J'ai participé à la préparation de certains petits ARN ovariens pour le séquençage, à certaines analyses bio-informatiques ainsi qu'aux expériences de dosage de pigments dans les yeux de drosophiles

Article II

Paramutation in Drosophila Requires Both Nuclear and Cytoplasmic Actors of the piRNA Pathway and Induces Cis-spreading of piRNA Production

Paramutation in *Drosophila* Requires Both Nuclear and Cytoplasmic Actors of the piRNA Pathway and Induces *Cis*-spreading of piRNA Production

Catherine Hermant,^{*,†} Antoine Boivin,^{*,†} Laure Teyssset,^{*,†} Valérie Delmarre,^{*,†} Amna Asif-Laidin,^{*,†}

Marius van den Beek,^{*,*,§} Christophe Antoniewski,^{*,*,§} and Stéphane Ronsseray^{*,†,1}

^{*}Sorbonne Universités, Université Pierre et Marie Curie - Paris 6, Institut de Biologie Paris-Seine, UMR7622, Laboratoire Biologie du Développement, F-75005, Paris, France, [†]CNRS, UMR7622, "Epigenetic Repression and Mobile DNA", F-75005, Paris, France,

[‡]CNRS, UMR7622, "Drosophila Genetics and Epigenetics", F-75005, Paris, France, and [§]CNRS, FR3631, Institut de Biologie Paris-Seine, ARTbio Bioinformatics Analysis Facility, F-75005, Paris, France

ABSTRACT Transposable element activity is repressed in the germline in animals by PIWI-interacting RNAs (piRNAs), a class of small RNAs produced by genomic loci mostly composed of TE sequences. The mechanism of induction of piRNA production by these loci is still enigmatic. We have shown that, in *Drosophila melanogaster*, a cluster of tandemly repeated *P-lacZ-white* transgenes can be activated for piRNA production by maternal inheritance of a cytoplasm containing homologous piRNAs. This activated state is stably transmitted over generations and allows *trans*-silencing of a homologous transgenic target in the female germline. Such an epigenetic conversion displays the functional characteristics of a paramutation, *i.e.*, a heritable epigenetic modification of one allele by the other. We report here that piRNA production and *trans*-silencing capacities of the paramutated cluster depend on the function of the *rhino*, *cutoff*, and *zucchini* genes involved in primary piRNA biogenesis in the germline, as well as on that of the *aubergine* gene implicated in the ping-pong piRNA amplification step. The 21-nt RNAs, which are produced by the paramutated cluster, in addition to 23- to 28-nt piRNAs are not necessary for paramutation to occur. Production of these 21-nt RNAs requires *Dicer-2* but also all the piRNA genes tested. Moreover, cytoplasmic transmission of piRNAs homologous to only a subregion of the transgenic locus can generate a strong paramutated locus that produces piRNAs along the whole length of the transgenes. Finally, we observed that maternally inherited transgenic small RNAs can also impact transgene expression in the soma. In conclusion, paramutation involves both nuclear (*Rhino*, *Cutoff*) and cytoplasmic (*Aubergine*, *Zucchini*) actors of the piRNA pathway. In addition, since it is observed between nonfully homologous loci located on different chromosomes, paramutation may play a crucial role in epigenome shaping in *Drosophila* natural populations.

KEYWORDS gene regulation; *trans*-generational epigenetics; noncoding small RNAs; mobile DNA; *Drosophila*

GENOMES must confront the presence of a large fraction of mobile DNA whose activity can result in severe deleterious effects on chromosome stability and gametogenesis. In the germline of animals, a system of genomic traps exists into which any transposable element (TE) can insert, thereby

generating loci that contain a catalog of potentially dangerous sequences that have to be repressed (Brennecke *et al.* 2007; Pane *et al.* 2011; Iwasaki *et al.* 2015). In the *Drosophila melanogaster* germline, most of these loci are transcribed in both directions (dual-strand clusters) and undergo noncanonical transcription and RNA processing (Mohn *et al.* 2014; Zhang *et al.* 2014). This results in production of noncoding small RNAs having the capacity to target the transcripts of the homologous, potentially active, TE copies scattered throughout the genome. These small RNAs are called PIWI-interacting RNAs (piRNAs) and repress TE activity at both the transcriptional and post-transcriptional levels (Sato and Siomi 2013; Weick and Miska 2014). piRNA biogenesis in the germline involves a nuclear and a cytoplasmic step. In the nucleus,

Copyright © 2015 by the Genetics Society of America

doi: 10.1534/genetics.115.180307

Manuscript received July 8, 2015; accepted for publication October 5, 2015; published Early Online October 13, 2015.

Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.180307/-/DC1.

Accession numbers: Small RNA sequences have been deposited at the ENA under accession numbers PRJEB11491.

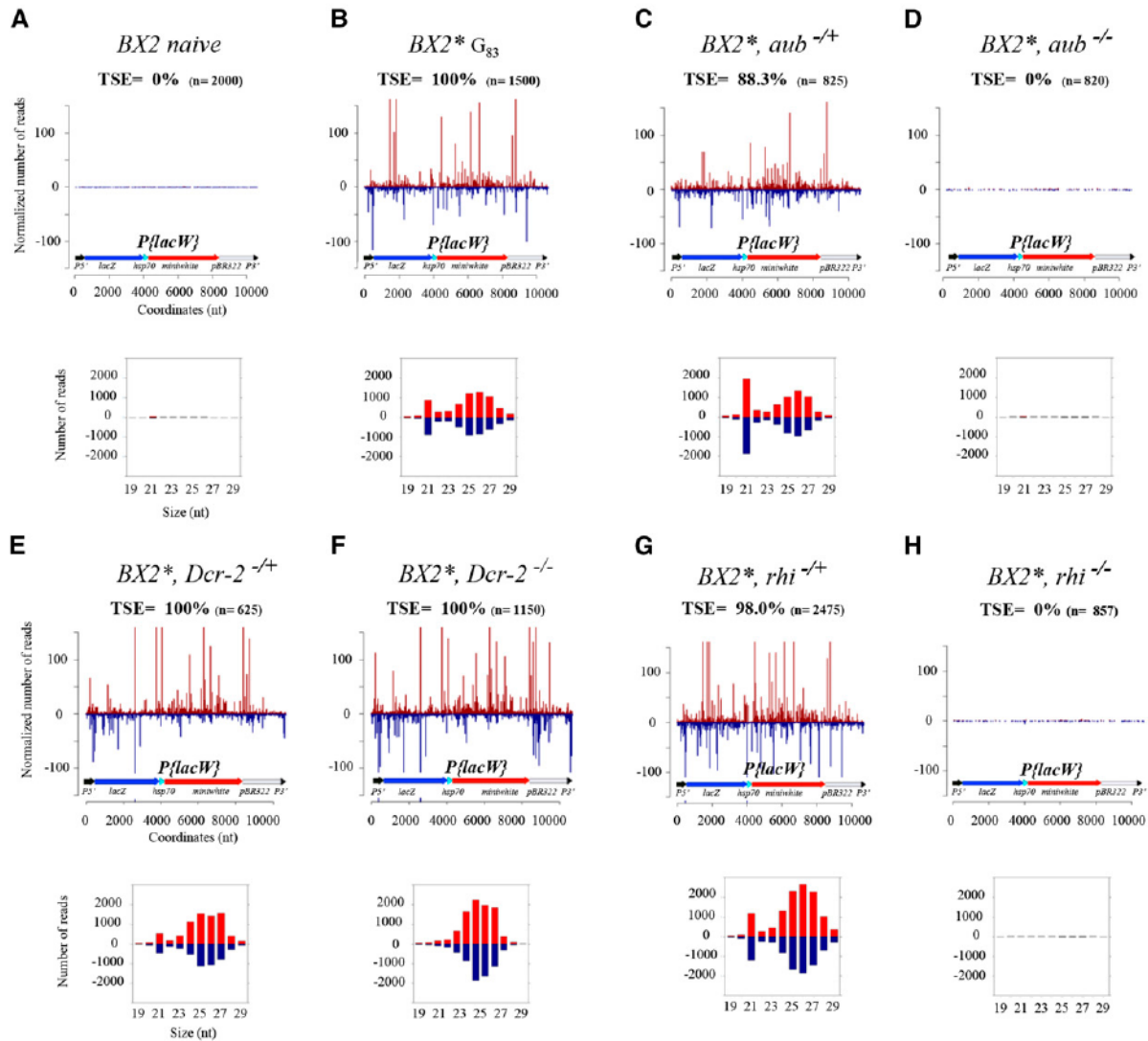
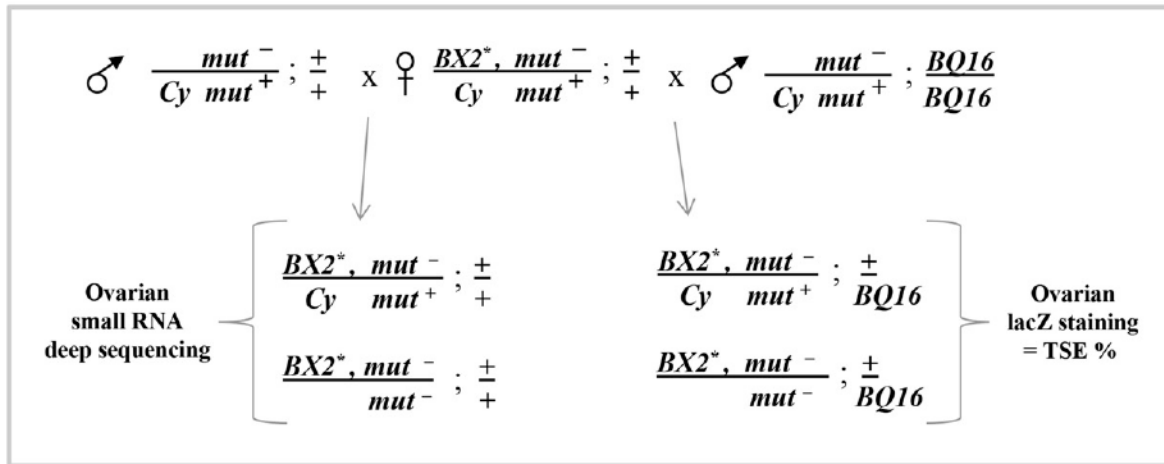
¹Corresponding author: "Epigenetic Repression and Mobile DNA," Institut de Biologie Paris-Seine, UMR7622, CNRS, University Pierre et Marie Curie, 9 Quai Saint-Bernard (Case 24), 75005, Paris, France. E-mail: stephane.ronsseray@upmc.fr

piRNA-producing cluster transcription requires in most of the cases the presence of the HP1 paralog Rhino associated with Deadlock and Cutoff on the locus, forming the so-called RDC complex (Klattenhoff *et al.* 2009; Pane *et al.* 2011; Mohn *et al.* 2014; Zhang *et al.* 2014). In the cytoplasm, transcripts produced by the piRNA locus are sliced in an optically dense region surrounding the nucleus, called the nuage, and small RNAs (23–28 nt) loaded on Piwi or Aubergine proteins are produced (primary piRNAs). Further, piRNAs loaded on Piwi enter the nucleus and target euchromatic TE copies to induce their transcriptional repression via heterochromatin formation, which involves HP1 (Wang and Elgin 2011; Sienski *et al.* 2012; Le Thomas *et al.* 2013; Rozhkov *et al.* 2013). In contrast, piRNAs loaded on Aubergine remain in the nuage and target homologous transcripts being exported from the nucleus, which are produced by both homologous TE copies and the piRNA locus. This produces secondary piRNAs loaded on Aubergine or Ago3, other PIWI proteins (Weick and Miska 2014; Iwasaki *et al.* 2015) and results in a piRNA amplification process called ping-pong amplification (Brennecke *et al.* 2007; Gunawardane *et al.* 2007). Finally, piRNA loaded on Piwi can also be produced downstream of the ping-pong amplification step by a slicing process that can spread on targeted RNA, increasing both piRNA quantity and diversity (Han *et al.* 2015; Mohn *et al.* 2015; Siomi and Siomi 2015). Upstream of this complex machinery, the presence of Rhino on the piRNA-producing locus is particularly important since it appears sufficient to promote processing of transcripts by the piRNA machinery (Klattenhoff *et al.* 2009; Zhang *et al.* 2014). How Rhino is addressed to a piRNA-producing locus is still unclear.

In *Drosophila*, in contrast to *Caenorhabditis elegans* (Ruby *et al.* 2006; Batista *et al.* 2008), the production of piRNAs by a piRNA locus in the germline does not appear to be only genetically determined, *i.e.*, no specific sequence motif or structure has been identified that is sufficient to promote piRNA production by DNA adjacent to this sequence. Conversely, using piRNA-producing loci that repress *P*-transposable elements (Ronsseray *et al.* 1996), it was shown that maternal transmission of piRNAs together with the piRNA locus can stimulate production of piRNAs by paternally inherited *P*-elements scattered through the genome (Brennecke *et al.* 2008). In addition, analysis of ageing effects on *I*-transposable element repression capacities showed that the amount of *I*-homologous piRNAs in adult ovaries is correlated to the amount of homologous piRNAs deposited in embryos (Grentzinger *et al.* 2012). Furthermore, it was shown that *de novo*, long-term activation of a piRNA locus can be achieved in *Drosophila* by maternal transmission of homologous piRNAs, without transmission of the initial piRNA donor locus. Under these conditions, there is emergence of a new autonomous and stable piRNA locus (de Vanssay *et al.* 2012). Indeed, a transgenic cluster of *P*-element deriving transgenes (called *BX2*), inert for repression and piRNA production, can be activated by crossing with females bearing a transgenic cluster (called *T-1*) that produces abundant piRNAs in both orienta-

tions homologous to the *BX2* cluster (de Vanssay *et al.* 2012). Once activated, this new piRNA cluster, called *BX2**, is stable over generations ($n > 100$) in absence of the inducer *T-1* locus and is able to produce abundant transgenic sense and antisense piRNAs. Cytoplasm of oocytes produced by these females can again *de novo* activate piRNA production by a paternally inherited inactive *BX2* cluster (de Vanssay *et al.* 2012). This recurrent epigenetic conversion process presents all the features of a “paramutation” process, a phenomenon previously described in plants (Brink 1956; Coe 1959) and recently in worms (Shirayama *et al.* 2012; Sapetschnig *et al.* 2015), and described as “an epigenetic interaction between two alleles of a locus, through which one allele induces a heritable modification of the other allele without modifying the DNA sequence” (Brink 1956; Chandler 2007). It was shown further that a *BX2** locus has a modified chromatin structure, showing an increase in H3K9 methylation with regard to the nonparamutated *BX2* locus (Le Thomas *et al.* 2014). This example of paramutation thus results in communication within a genome between homologous clusters allowing one cluster to transfer piRNA production capacity to a previously inert cluster. Paramutation may play an important role in establishment of TE repression following a genetic invasion since master sites of repression could activate piRNA secondary sites thus reinforcing their repressive capacity. Investigating the mechanism of *de novo* activation of piRNA-producing loci is therefore important for understanding aspects of epigenome shaping in general.

We have therefore analyzed the functional requirements and properties of a *BX2** paramutated locus. This locus allows the investigation of the autonomous properties of a piRNA-producing locus, *i.e.*, independently of the influence of homologous sequences present in the genome that could interfere with the process. The *BX2** locus was previously shown to produce, not only abundant transgenic piRNAs, but also 21-nt RNAs inferred to be small-interfering RNAs (siRNAs), but the precise nature of these 21-nt RNAs and their functional role in silencing capacities were not elucidated. Here, we show first that *BX2** silencing and piRNA production capacities depend on the *rhino* and *cutoff* genes, involved in the noncanonical transcription of piRNA-producing loci, and on the *aubergine* and *zucchini* genes, involved in piRNA cytoplasmic processing. In addition, transgenic 21-nt RNA production depends on *Dcr-2*, which characterizes them as siRNAs but they are not necessary for paramutation to occur. Surprisingly, production of these siRNAs depends also on the piRNA pathway. Moreover, paramutation can occur between only partially homologous loci within a genome and is accompanied by a rapid spreading of the capacity to produce piRNAs to the nonhomologous sequences within the paramutated locus. Finally, an impact of maternally inherited small RNAs linked to paramutation is also observed in somatic cells. These findings allow us to characterize more precisely the molecular mechanism of paramutation in animals and to propose that it is a key factor in epigenome shaping in natural populations.



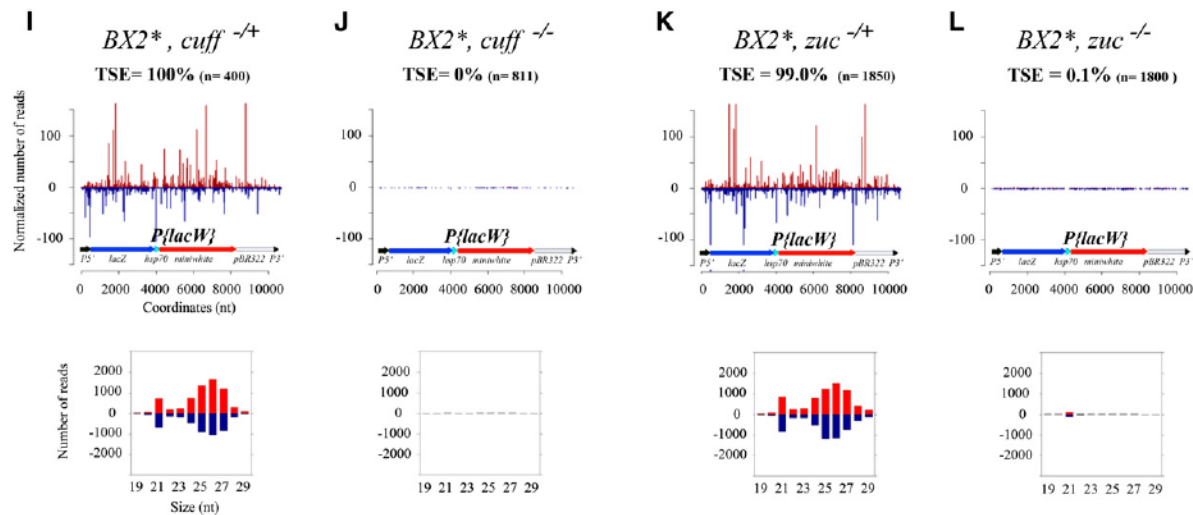


Figure 1 Effect of mutations affecting the piRNA or siRNA pathways on *trans*-silencing and small RNA production capacities of a *BX2** cluster. Upper box: Mating scheme used to analyze the effect of mutations (*mut*) affecting *aubergine*, *Dicer-2*, *rhino*, *cutoff*, and *zucchini* on *BX2** ovarian small RNA production and *trans*-silencing capacities. As the *BX2* cluster, all genes are located on chromosome 2. (Left) Heterozygous *BX2* aub*⁻, *BX2* Dcr-2*⁻, *BX2* rhi*⁻, *BX2* cuff*⁻, or *BX2* zuc*⁻ females were crossed with heterozygous males carrying a mutant allele of the same gene to generate either heterozygous control or loss-of-function females for deep sequencing of ovarian small RNAs. (Right) The same heterozygous females were crossed with heterozygous males carrying a *P-lacZ* target transgene (*BQ16*) to measure the effect of the mutation tested on *BX2** *trans*-silencing capacities (TSE %). TSE was quantified by determining the percentage of egg chambers with no *lacZ* expression in the germline. In the genotypes presented, maternally inherited chromosomes (chromosomes 2 and 3) are indicated above the bar. Cy is a balancer chromosome carrying a dominant phenotypic marker. (A–L) In each case, the genotype tested is indicated, and the percentage of TSE is given below the genotype with the total number of egg chambers assayed in parentheses. Nonparamutated *BX2* ("naive") and *BX2** at generation 83 were analyzed as controls. Histograms show the length distributions of ovarian small RNAs matching the *P(lacW)* locus. Positive and negative values correspond to sense and antisense reads, respectively. Plots show the abundance of 19- to 29-nt small RNAs matching the *P(lacW)* locus. Analysis of the effect of *aub*, *rhi*, *cuff*, and *zuc* mutations shows that *BX2** silencing capacities, ovarian *P(lacW)* 23- to 28-nt piRNA, but also 21-nt RNA production depend on the piRNA pathway. *Dcr-2* loss of function strongly affects production of *P(lacW)* homologous 21-nt RNAs without affecting that of 23- to 28-nt RNAs or *trans*-silencing capacities. *P(lacW)* 21 nt are not necessary for the maintenance of the paramutated state and their biogenesis depends on both *Dcr-2* and the piRNA pathway. TSE results for *Dcr-2* mutants (E and F) are reprinted from de Vanssay *et al.* (2012).

Materials and Methods

Experimental conditions

All crosses were performed at 25° and involved three to five couples in most cases. *lacZ* expression assays were carried out using X-gal overnight staining as described in Lemaitre *et al.* (1993), except that ovaries were fixed for 8 min. *Trans*-silencing was quantified as previously described by determining the percentage of egg chambers with no *lacZ* expression in the germline (Josse *et al.* 2007). It is referred to as the *trans*-silencing effect percentage (TSE %).

Dissection and RNA extraction

Total RNAs were extracted using TRIzol (Life Technologies) as described in the reagent manual (http://tools.lifetechnologies.com/content/sfs/manuals/trizol_reagent.pdf). For the RNA precipitation step, 100% ethanol was used instead of isopropanol. For ovarian small RNAs, 25 pairs of ovaries were manually dissected for each sample in PBS 1×.

Pigment dosage

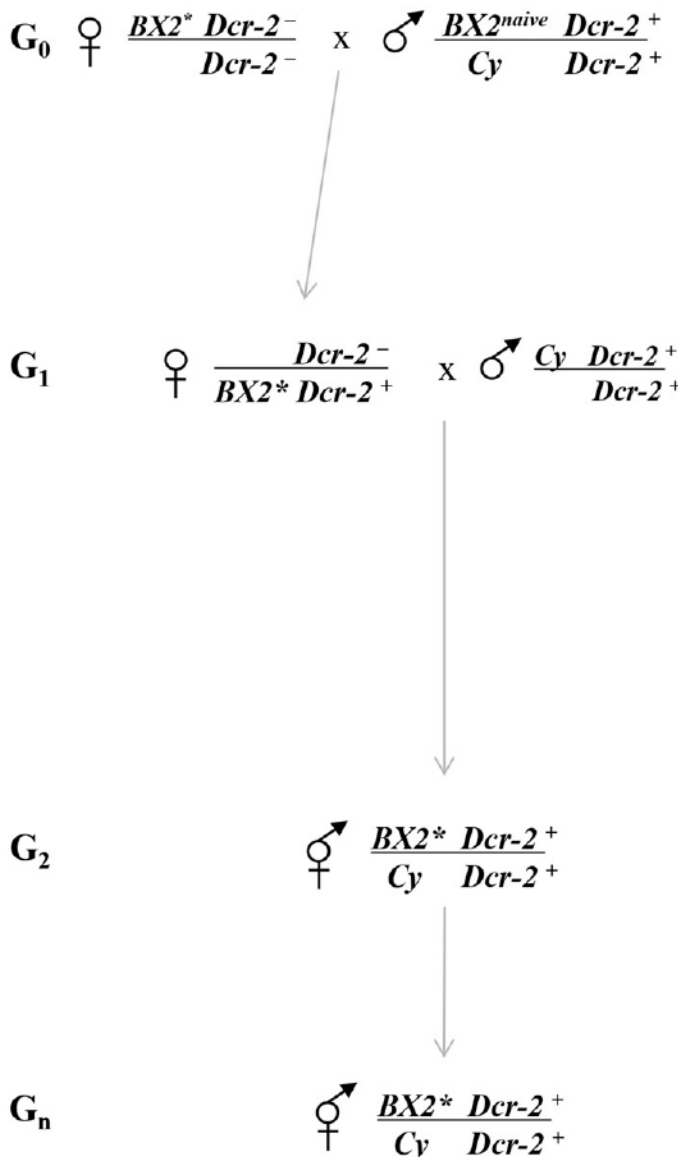
Head pigments were quantified as described in Weiler and Wakimoto (2002). For each genotype, three replicates were

measured. For each replicate, five heads of 4- to 6-day-old females were manually dissected for the dosage. Absorbance was measured at 480 nm.

Transgenes and strains

P-lacZ fusion enhancer-trap transgenes: *P-1152* and *BQ16*, contain an in-frame translational fusion of the *Escherichia coli lacZ* gene to the second exon of the *P*-transposase gene and a *rosy* transformation marker (O'Kane and Gehring 1987). The *P-1152* insertion was mapped to the telomere of the X chromosome (cytological site 1A) and consists of two *P-lacZ* insertions in the same telomeric-associated sequence unit (TAS) (Karpen and Spradling 1992) and in the same orientation (Josse *et al.* 2007). *P-1152* is homozygous viable and fertile. *BQ16* is located at 64C in euchromatin of the third chromosome and is homozygous viable and fertile. *P-1152* shows no *lacZ* expression in the ovary; *BQ16* is strongly expressed in the nurse cells and in the oocyte (Josse *et al.* 2007). *RS3* is a *P-FRT-white* transgene (FBtp0001534). It is inserted in the TAS of the 3R chromosomal arm (cytological site 100E3). It is homozygous viable and fertile (DGRC, Kyoto no. 123282). *P-1152* and *RS3* induce a strong TSE and produce abundant ovarian transgenic piRNAs

A



B

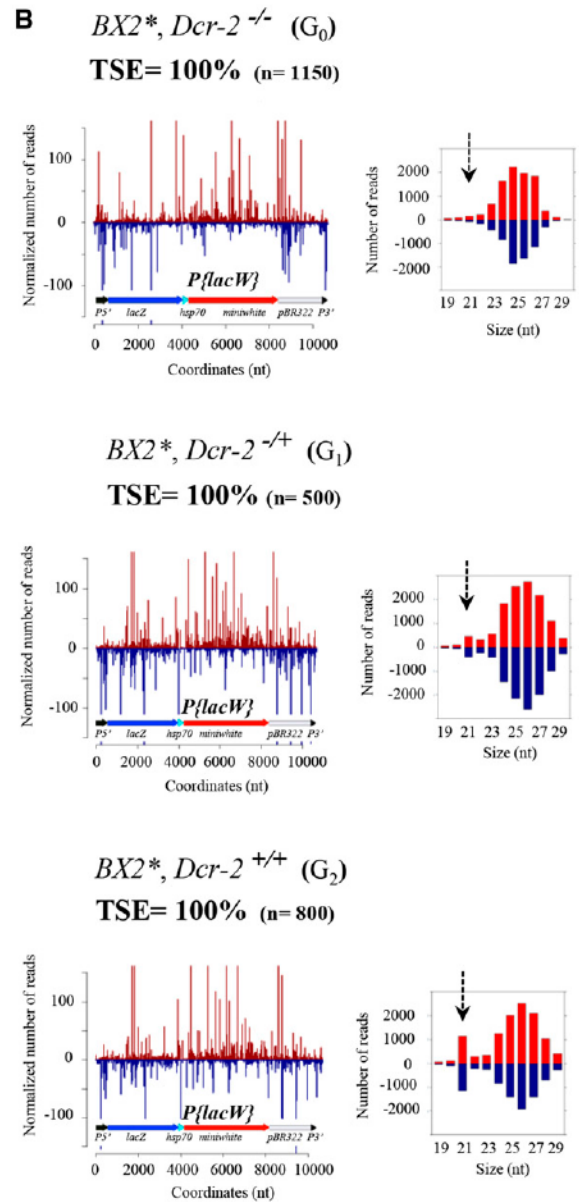


Figure 2 Paramutation by a cytoplasm devoid of siRNAs. (A) *BX2** *Dcr-2*^{-/-} females were crossed to *BX2* *Dcr-2*^{+/+} (naive) males to generate *G*₁ females having paternally inherited a *BX2* locus and maternally inherited piRNAs but no 21-nt siRNAs homologous to the *P*[*lacW*] cluster. In addition, these *G*₁ females paternally inherited a *Dcr-2*⁺ allele. These *G*₁ females were crossed with *Dcr-2*⁺ males to establish a *BX2** *Dcr-2*^{+/+}/*Dcr-2*^{+/+} line in which no maternal *P*[*lacW*] siRNAs have been initially introduced in *G*₀. *Cy* is a balancer chromosome carrying a dominant phenotypic marker. Maternal chromosomal complement is written above the bar. (B) At generations 0, 1, and 2, deep sequencing of ovarian small RNAs was performed and silencing capacity was controlled in parallel by crossing females with males carrying the *BQ16* target transgene and scoring TSE in female progeny. The genotype tested is indicated and TSE percentages and ovarian 19- to 29-nt RNAs are presented as in Figure 1. As expected, *G*₁ and *G*₂ females show complete silencing capacities. The same was true for successive *G*_n generations (*G*₆, TSE = 100%, *n* = 450; *G*₁₀, TSE = 100%, *n* = 1100). Despite the lack of maternal *P*[*lacW*] siRNA maternal transmission in *G*₀, such ovarian small RNAs appear already in *G*₁ females and increase in *G*₂ in *BX2** *Dcr-2*^{+/+} females (dashed arrow). Therefore, maternal inheritance of *P*[*lacW*] homologous piRNAs, but no siRNAs, results in a strong and stable paramutated *BX2** locus and in immediate production of both *P*[*lacW*] piRNAs and siRNAs in females carrying a *Dcr2*⁺ allele.

(Roche and Rio 1998; Josse *et al.* 2007; de Vanssay *et al.* 2012; Dufourt *et al.* 2014) and this study (see Figure 3A). The *Lk-P*(1A) line carries two full-length *P*-elements inserted in TAS at the *X*-chromosome telomere (Ronsseray *et al.* 1991).

These telomeric *P*-insertions derive from a chromosome present in a natural population (Biemont *et al.* 1990) and induce very strong repression of *P*-element transposition and *P*-induced hybrid dysgenesis (Ronsseray *et al.* 1991, 1996).

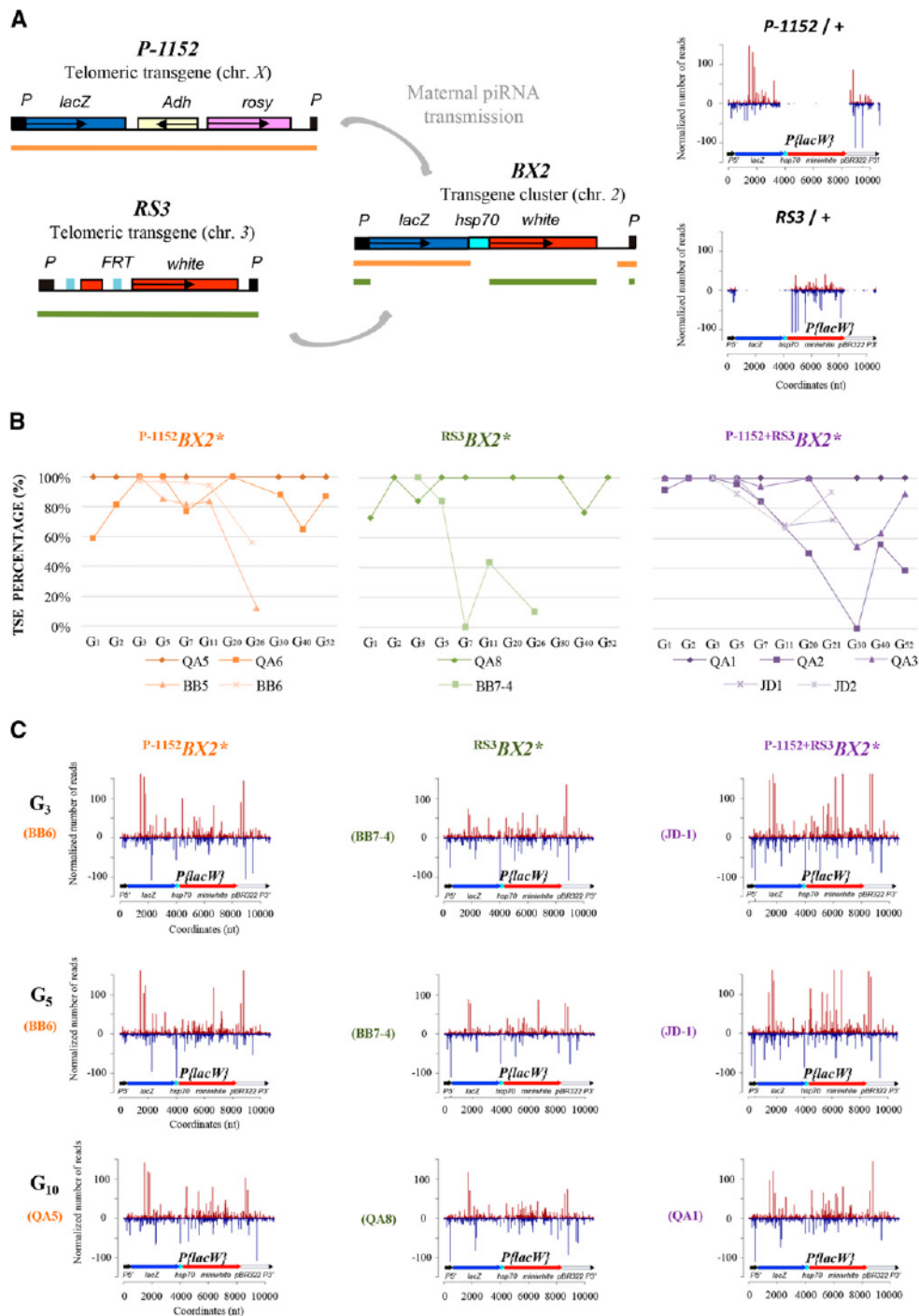


Figure 3 Paramutation of a *BX2* locus by partially homologous piRNAs loci. *BX2* males carrying a *P{lacW}* naive cluster were crossed to females carrying partially homologous *P*-transgenes inserted in telomeric piRNA-producing loci at the heterozygous state. Female progeny were recovered that carry the *BX2* locus but did not carry the telomeric transgenes. These *G*₁ females inherited a cytoplasm carrying piRNA homologous to the telomeric transgenes, which therefore cover only partially the *BX2* *P{lacW}* transgene sequence. Lines were established and the *trans*-silencing capacities of these putative *BX2** lines were tested in subsequent generations, as in Figure 1 and Figure 2. In addition, deep sequencing of ovarian small RNAs was performed in *G*₃, *G*₅, and *G*₁₀. (A, left and middle) Structure of the *P-1152* and *RS3* telomeric *P*-transgenes used and of the *BX2* *P{lacW}* transgenes. Colored thick lines below the transgenes indicate the sequences shared by the *P{lacW}* and telomeric transgenes. The *P-1152* line carries two *P-lacZ-rosey* (*P{larB}*) transgenes inserted in the telomeric associated sequences (TAS) of the X chromosome. *RS3* carries a *P-FRT-white* transgene in the TAS of the 3R chromosomal arm. These two lines induce a strong TSE. (A, right) Ovarian piRNA production in females that carry a maternally inherited *P-1152* or *RS3* locus at the heterozygous state (*G*₁ females from crosses between *P-1152* and *RS3* females with males devoid of transgene) is mapped on the *P{lacW}* transgene. (B) Three types of replicate *BX2** sublines were generated: Sublines that inherited in *G*₁ (1) a cytoplasm from females heterozygous for the *P-1152* locus (*P-1152**BX2* sublines, in orange); (2) a cytoplasm from females heterozygous for the *RS3* locus (*RS3**BX2* sublines, in green); and (3) a cytoplasm from females heterozygous for both the *P-1152* and *RS3* loci (*P-1152+RS3**BX2* sublines, in purple). *Trans*-silencing capacities of various *BX2** lines are shown with regard to generations (TSE %, *n* > 500 in all assays). Names of replicate sublines are given below the graphs. (C) Deep sequencing of ovarian small

RNAs in various *BX2** sublines at generations 3, 5, and 10. Plots show the abundance of 19- to 29-nt small RNAs matching *P{lacW}*. TSE assays show that silencing capacities vary among replicate *BX2** lines but strong and stable paramutated lines can be recovered with the three types of *G*₀ cytoplasmic inheritance (B). Ovarian small RNAs were analyzed for a subset of the *BX2** lines showing strong silencing capacities. The name of the line is indicated on the left of the graph. Deep sequencing analysis shows that piRNAs corresponding to the whole length of *P{lacW}* can be produced as early as generation 3 (C), showing that piRNA production by a paramutated locus can rapidly extend to the part of the locus that did not originally receive maternally inherited piRNAs.

Lk-P(1A) females produce abundant *P*-homologous ovarian piRNAs (Brennecke *et al.* 2008). The repressive capacity of the *Lk-P(1A)* line was verified immediately prior to the experiments reported in Supporting Information, Figure S6 and a very strong repression capacity was found (data not shown).

***P{lacW}* clusters:** The *BX2* line carries seven *P-lacZ-white* transgenes, including at least one defective copy, inserted in tandem and in the same orientation at cytological site 50C on the second chromosome (Dorer and Henikoff 1994). The transgene insertion site is located near the

mRpl53 gene in an *Ago1* intron (de Vanssay *et al.* 2012). This site is not a piRNA-producing locus, as observed for instance in the deep-sequencing dataset from *P-1152* or *w¹¹¹⁸* ovaries (data not shown). The *P-lacZ-white* construct contains the *P-lacZ* translational fusion and is marked by the mini-*white* gene (*P{lacW}*, FBtp0000204, 10,691 bp). The *T-1* line derives from X-ray treatment of *BX2* (Dorer and Henikoff 1997). *T-1* has chromosomal rearrangements including translocations between the second and third chromosomes. *DX1* was generated during the same *P-lacZ-white* transposase-mediated remobilization at 50C than *BX2* and carries six *P-lacZ-white* transgenes, one of which is in opposite orientation to the others, at the same genomic site as *BX2* (Dorer and Henikoff 1994, 1997). After overnight staining, weak *lacZ* expression is detected in the follicle cells of *BX2*, *T-1*, and *DX1* female ovaries, presumably because of a position effect at 50C, but no staining is observed in the germline (data not shown). The three *P{lacW}* clusters show variegated repression of *white* in the eyes due to repeat induced gene silencing (RIGS) (Dorer and Henikoff 1994, 1997). The extent of variegated repression of *white* is different between the clusters (*T-1* >> *DX1* > *BX2*). The *BX2* cluster was not used to assay the somatic phenotypic impact of piRNA maternal inheritance since RIGS-induced repression is too low in this case.

Lines carrying transgenes have M genetic backgrounds (devoid of *P*-transposable elements), as do the multimarked balancer stocks used in genetic experiments. The *w¹¹¹⁸* and Canton^y lines were used as controls completely devoid of any *P*-element or transgene. Crosses performed with *BX2*, *T-1*, or *DX1* were performed with females carrying the cluster at the heterozygous state because of the sterility (*BX2*, *DX1*) and lethality (*T-1*) induced by transgene clusters.

Two strong hypomorphic mutant alleles of *aubergine* (*aub*) induced by EMS were used. Both of them are homozygous female sterile and TSE was previously shown to be abolished by a heteroallelic combination of these alleles (Josse *et al.* 2007). *aub^{QC42}* comes from the Bloomington Stock Center (stock no. 4968) and has not been characterized at the molecular level (Schupbach and Wieschaus 1991). *aub^{N11}* has a 154-bp deletion, resulting in a frameshift that is predicted to add 16 novel amino acids after residue 740. *Dicer-2^{L811fsx}* (*Dcr-2^{L811fsx}*) is a loss-of-function allele induced by EMS, which has a sequence variant at residue 811 resulting in a stop codon (Lee *et al.* 2004). It is homozygous viable and fertile. *rhino⁰²⁰⁸⁶* (*rhi²*) results from a *P{PZ}* insertion (*P-lacZ-rosy*) at nucleotide 267. The *P{PZ}* transgene is oriented in the opposite direction to that of *rhi* transcription (Volpe *et al.* 2001) (Bloomington no. 12226). *rhi^{KG00910}* (*rhi^{KG}*) is due to the insertion of a *P{SUPor-P}* transgene, which carries the *yellow* and *white* transformation markers (Gene Disruption Project 2001, Bloomington no. 13161). *cutoff^{WM25}* (*cuff^{WM25}*) and *cuff^{QQ37}* alleles were isolated from an EMS screen (Schupbach and Wieschaus 1991). *cuff^{WM25}* corresponds to the replacement of the first methionine by a lysine and *cuff^{QQ37}* has a T to A substitution at position 786 that could result in a splicing alteration (data not shown). *zucchini*

(*zuc*) alleles were isolated from an EMS screen (Schupbach and Wieschaus 1991). *zuc^{HM27}* contains a stop codon at residue 5 and *zuc^{SG63}* a substitution of histidine 169 with a tyrosine in the conserved HKD domain presumably involved in nuclease activity. TSE was previously shown to be abolished by a heteroallelic combination of these alleles (Todeschini *et al.* 2010). The crosses performed to generate heteroallelic mutant *BX2** females are indicated in Table S1. When information was available, the most severe allele was introduced maternally. Lines carrying mutations of *zuc* and *cuff* were kindly provided by Attilio Pane and Trudi Schupbach and *aub^{N11}* was kindly provided by Paul Macdonald.

All the alleles described above are located on the second chromosome and are maintained over a Cy balancer chromosome. Additional information about mutants and stocks are available at FlyBase: <http://flybase.bio.indiana.edu/>.

Deep sequencing analyses

A small RNA fraction, from 18 nt to 30 nt in length, was obtained by separating it on a denaturing polyacrylamide gel from total RNA extracted from dissected ovaries. This fraction was used to generate multiplexed libraries with Illumina TruSeq Small RNA Library preparation kits (RS-200-0012, RS200-0024, RS-200-036, or RS-200-048) at Fasteris (<http://www.fasteris.com>). A house protocol based on TruSeq, which reduces 2S RNA (30 nt) contamination in the final library, was performed (except for libraries whose GRH number is below no. 40). Libraries were sequenced using Illumina HiSeq2500 and 2500. Sequence reads in fastq format were trimmed from the adapter sequence 5'-CTGTAGGCAC CATCAATCGTA-3' (GRH12, GRH13, GRH14, GRH17) or 5'-TGGAAATCTCGGGTGCCAAG-3' (other samples) and matched to the *D. melanogaster* genome release 5.49 using Bowtie (Langmead *et al.* 2009) and to the sequences of the *P*-element constructs *P{lacW}* (FlyBase ID FBtp0000204) as indexed reference. Only 19- to 29-nt reads matching the reference sequences with 0 or 1 mismatch were retained for subsequent analysis. For global annotation of the libraries (Table S1), we used the release 5.49 of fasta reference files available in FlyBase, including transposon sequences (dmel-all-transposon_r5.49.fasta) and the release 20 of miRNA sequences from miRBase (www.mirbase.org).

Sequence length distributions, small RNA mapping, and small RNA overlap signatures were generated from bowtie alignments using Python and R (www.r-project.org/) scripts, which were wrapped and run in a Galaxy instance publicly available at <http://mississippi.fr>. Tools and workflows used in this study may be downloaded from this Galaxy instance. For library comparisons, read counts were normalized (Table S1) to the total number of small RNAs that matched the *D. melanogaster* genome (release 5.49) and did not correspond to abundant cellular RNAs (tRNAs and miscRNAs). For small RNA mapping (Figure 1, Figure 2, and Figure 3 and Figure S1, Figure S3, and Figure S6), we took into account only RNA reads that uniquely aligned to *P{lacW}* or the 42AB locus.

Distributions of piRNA overlaps (ping-pong signatures, Figure S2 and Figure S5) were computed as first described in Klattenhoff *et al.* (2009) and detailed in Antoniewski (2014). Thus, for each sequencing dataset, we collected all the 23- to 28-nt RNA reads matching *P{lacW}* or the 42AB locus whose 5' ends overlapped with another 23- to 28-nt RNA read on the opposite strand. Then, for each possible overlap of 1–28 nt, the number of read pairs was counted. Distributions of siRNA overlaps (Figure S3) were computed using a similar procedure, except that 20- to 21-nt RNA reads were collected instead of the 23- to 28-nt RNA reads. The distributions of piRNA/siRNAs overlaps (Figure S3) were computed by collecting separately the 20- to 21-nt and 23- to 28-nt RNA reads matching *P{lacW}* or the 42AB locus and counting for each possible overlap of 1–21 nt the number of read pairs across these two distinct read datasets. To plot the overlap signatures, a z-score was calculated by computing, for each overlap of 1 to *i* nucleotides, the number *O(i)* of read pairs and converting it, using the formula $z(i) = (O(i) - \text{mean}(O)) / \text{standard deviation}(O)$.

Data availability

Small RNA sequences have been deposited at the European Nucleotide Archive (ENA) (<http://www.ebi.ac.uk/ena>) under accession number PRJEB11491.

Results

Effect of mutation affecting the piRNA or siRNA pathway on BX2* small RNA production and silencing capacities

Paramutated BX2* females were previously shown to produce 23- to 28-nt and 21-nt ovarian sense and antisense small RNAs homologous to the cluster of *P-lacZ-white* transgenes (called *P{lacW}*) and to induce complete silencing of homologous transgene expression in the female germline (de Vanssay *et al.* 2012). This homology-dependent silencing, which reflects piRNA functionality, is quantified using the *trans*-silencing effect (TSE) assay (Josse *et al.* 2007) in which the percentage of repressed egg chambers is measured. The 23- to 28-nt, but not 21-nt, RNAs presented the molecular signature of the ping-pong amplification step of the piRNA pathway. This strongly suggests that production of the 23- to 28-nt RNAs is dependent on the piRNA pathway, whereas that of the 21-nt RNAs would depend on the siRNA pathway. Moreover, the TSE capacities of the BX2* paramutated locus were shown to be completely impaired by loss of function of *aubergine* (*aub*) involved in the piRNA ping-pong step, whereas they were insensitive to loss of function of *Dicer-2* (*Dcr-2*) involved in the siRNA pathway (de Vanssay *et al.* 2012). This suggested that the 21 nt would not be necessary for BX2* silencing capacities.

We thus performed deep sequencing of ovarian small RNAs of mutant females for *aub* and *Dcr-2* and extended the analysis to *rhino* (*rhi*) and *cutoff* (*cuff*), which are nuclear actors

involved in piRNA loci noncanonical transcription (Mohn *et al.* 2014; Zhang *et al.* 2014), and to *zucchini* (*zuc*), which is involved in the cleavage of piRNA precursors and secondary piRNAs in the cytoplasm (Ipsaro *et al.* 2012; Nishimasu *et al.* 2012; Voigt *et al.* 2012; Han *et al.* 2015; Mohn *et al.* 2015). All these genes are located, as BX2 is, on chromosome 2. For each gene tested, a recombinant chromosome carrying a BX2* locus and a mutated allele was generated. Females heterozygous for this recombinant chromosome were crossed with heterozygous males, carrying one mutant allele of the corresponding tested gene, to generate control heterozygote, loss-of-function heteroallelic (*aub*, *rhi*, *cuff*, *zuc*) and homozygote (*Dcr-2*) females for deep sequencing of ovarian small RNAs (Figure 1, upper box, and Table S1). In addition, females heterozygous for recombinant chromosomes were crossed with heterozygous males carrying a corresponding mutated allele and a target *P-lacZ* transgene expressed in the female germline to investigate the effect of the mutation tested on the *trans*-silencing capacities of BX2* (Figure 1, upper box). As controls, we analyzed small RNAs from ovaries of BX2* females at generation 83 (G₈₃) after the paramutation process and of BX2 nonparamutated females (called “BX2 naive”). BX2* lines maintained at 25° show a very strong stability through generations, as tested every five generations, showing a complete repression capacities (for example G₈₃: TSE = 100%, *n* = 1500; G₁₁₅: TSE = 100%, *n* = 1150). Figure 1 shows control females together with heterozygous and homozygous females for loss-of-function mutations for each gene tested, in terms of their capacity for *trans*-silencing (TSE %) and for production of ovarian small RNAs homologous to the *P{lacW}* transgene. As expected, the BX2 naive cluster showed no silencing capacities (TSE = 0%, *n* = 2000) and did not produce significant amounts of *P{lacW}* homologous 23- to 28-nt or 21-nt RNAs (Figure 1A) whereas the BX2* paramutated cluster continued to induce complete *trans*-silencing and to produce high levels of *P{lacW}* homologous 23- to 28-nt and 21-nt RNAs 83 generations after the paramutation process (Figure 1B). *aub* loss of function resulted in the complete loss of 23- to 28-nt RNAs homologous to the *P{lacW}* transgene (Figure 1D), when compared to heterozygous *aub* females (Figure 1C). Surprisingly, *aub* loss of function also affected production of the 21-nt RNAs by the *P{lacW}* cluster (Figure 1D). This did not result from a genomic background effect on the siRNA pathway in *aub* mutant females since the production of endogenous siRNAs at the *esi-1* locus was not affected (Figure S4, D). *Dcr-2* loss of function resulted in a strong decrease of 21-nt RNAs homologous to the *P{lacW}* transgene but the production of corresponding 23- to 28-nt RNAs was not affected, nor their ping-pong signature (Figure 1F and Figure S2J). Analysis of mutations affecting *rhi*, *cuff* or *zuc* produced similar results to that of *aub* mutations. TSE percentages showed that *rhi*, *cuff*, or *zuc* loss of function resulted in loss of BX2* silencing capacities (Figure 1, H, J, and L), whereas no effect was found with *rhi*, *cuff*, and *zuc* heterozygotes (Figure 1, G, I, and K). *rhi*, *cuff*, and *zuc* loss of function also

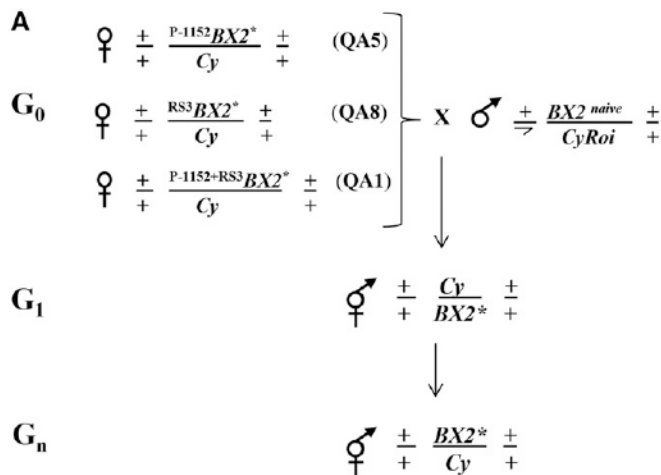


Figure 4 Paramutagenicity of a partially homologous paramutation. $BX2^*$ lines having, at G₀, maternally inherited piRNA homologous to a part of the $BX2$ $P\{lacW\}$ transgene sequence, and showing strong silencing capacities at the 20th generation (Figure 3), were tested for their ability to *de novo* paramutate a naive $BX2$ locus. (A) The name of the line is indicated beside the G₀ female genotype. Maternal chromosomal complement is written above the bar. Cy and $CyRoi$ are balancer chromosomes devoid of transgenes and carrying different dominant markers. These $P-1152 BX2^*$, $RS3 BX2^*$, and $P-1152+RS3 BX2^*$ females were crossed with males carrying a $BX2$ naive cluster and G₁ progeny having paternally, but not maternally, inherited a $BX2$ cluster were crossed to establish lines. These lines carry a second order paramutated locus. (B) These lines were subsequently tested for *trans*-silencing capacities at generations 1, 4, and 36. The TSE percentage is indicated with the number of egg chambers counted in parentheses. Strong silencing capacities are observed over generations showing that $P-1152 BX2^*$, $RS3 BX2^*$, and $P-1152+RS3 BX2^*$ are strongly paramutagenic.

resulted in complete loss of ovarian $P\{lacW\}$ piRNA production together with an almost complete loss of 21-nt RNAs homologous to the transgene (Figure 1, H, J, and L). As expected, these mutants exhibited no effect on the produc-

tion of endogenous siRNAs at the *esi-1* locus (Figure S4) but production of piRNAs by the 42AB typical dual-strand piRNA locus was affected (Figure S1). In conclusion, $BX2^*$ silencing capacities and $P\{lacW\}$ homologous piRNA production depend on all the piRNA pathway genes tested, including nuclear actors, and do not depend on *Dcr-2*. Interestingly, *Dcr-2*-dependent 21-nt RNAs' production by the cluster also depends on all the piRNA pathway genes tested. $BX2^*$ is thus a typical dual-strand piRNA cluster, requiring the primary and secondary piRNA pathways, which allows the study of the properties of a piRNA cluster, independently from any influence of endogenous homologous sequences in *trans*.

Production of ovarian transgenic siRNAs does not require siRNA maternal inheritance

Our previous work (de Vanssay *et al.* 2012), combined with the results in Figure 1F, shows that $BX2^*$ silencing properties and production of piRNAs homologous to the $P\{lacW\}$ transgene do not require maternal inheritance of siRNAs. Results in Figure 1 show that presence of ovarian $P\{lacW\}$ 21-nt siRNAs depends, in addition to *Dcr-2*, on the function of all piRNA pathway genes tested. We asked next if production of $P\{lacW\}$ ovarian siRNAs requires maternal inheritance of such siRNAs. We thus generated $BX2^*$ females carrying a paternally inherited *Dicer-2* wild-type allele and maternally inherited $P\{lacW\}$ homologous piRNAs but no siRNAs, because their G₀ mothers were homozygous for *Dcr-2* mutations (Figure 2). As expected (de Vanssay *et al.* 2012), these G₁ *Dcr-2* heterozygous females showed complete silencing capacities. Deep sequencing shows that, in addition to ovarian $P\{lacW\}$ piRNAs, they produce $P\{lacW\}$ siRNAs. Further, these G₁ females were crossed with *Dcr-2*^{+/+} males and increased ovarian $P\{lacW\}$ siRNA levels were found in $BX2^*$ G₂ females having recovered a *Dcr-2*^{+/+} genotype. Therefore, production of ovarian $P\{lacW\}$ siRNAs by females having a functional siRNA pathway does not require homologous siRNA inheritance and can be induced by inheritance of solely homologous piRNAs.

Paramutation of a locus only partially covered by maternally inherited piRNAs can be stable and is accompanied by cis-spreading of piRNA production capacity

Targeting of transgenes or natural transposable elements by piRNAs can lead to spreading of piRNA production to sequences adjacent to the targeted elements (Muerdter *et al.* 2011; Olovnikov *et al.* 2013; Shpiz *et al.* 2014). We tested if spreading could be observed for a silencer locus induced by paramutation. Telomeric transgenes inserted in subtelomeric heterochromatin can show strong *trans*-silencing capacities (Roche and Rio 1998; Ronsseray *et al.* 2003; Josse *et al.* 2007; Josse *et al.* 2008; Dufourt *et al.* 2014) and produce abundant piRNAs in the germline (Muerdter *et al.* 2011; de Vanssay *et al.* 2012). These transgenes have a different structure than the $P\{lacW\}$ transgene of $BX2$ but share some common sequences (de Vanssay *et al.* 2012). It was therefore possible to

produce females, having paternally inherited a *BX2* naive locus and maternally inherited cytoplasm containing piRNAs homologous to only a subset of the *P{lacW}* transgene structure. Two telomeric transgenic loci were used (Figure 3). The *P-1152* line carries two copies of the *P-lacZ-rosy* transgene (*P{larB}*) in the telomeric associated sequences (TAS) of the *X* chromosome (Josse *et al.* 2007). The *P-1152* telomeric locus was the canonical silencer locus used in previous *trans*-silencing effect studies (Roche and Rio 1998; Josse *et al.* 2007, 2008; Todeschini *et al.* 2010; Pöyhönen *et al.* 2012). *RS3* carries a *P-FRT-white* transgene in the TAS of the third chromosome (3R arm). It also induces a strong TSE (Dufourt *et al.* 2014). Thus, *BX2 P{lacW}* transgenes shares *P*-element and *lacZ* sequences with *P-1152* and shares *P*-element and *white* sequences with *RS3*. In both cases, a part of the *P{lacW}* sequence is missing in the telomeric transgenic construct (see Figure 3A). In order to transmit the cytoplasm, but not the telomeric transgenes, females heterozygous for these telomeric transgenes were generated. We further crossed *BX2* males, from a nonparamutated line, with females heterozygous for each or both telomeric transgenes. *G*₁ females having inherited the *BX2* locus, but not the telomeric transgenes, were recovered and potentially paramutated *BX2* lines were established and analyzed for *trans*-silencing and ovarian piRNA production capacities over generations. Figure 3B shows that *G*₁ females that carried a paternally inherited *BX2* cluster and recovered cytoplasm containing piRNAs produced by partially homologous transgenes, can show strong or complete silencing capacities. However, the efficiency was not as strong as that of paramutation experiments in which the *P{lacW}* transgenes of the *BX2* paramutated locus inherit piRNAs homologous to the entire transgene length, which always result in a complete and stable paramutation (de Vanssay *et al.* 2012 and data not shown). For the *BX2** lines activated by piRNAs homologous to the *P-1152* transgenes (called *P-1152BX2**), two lines (QA5 and QA6) showed very strong silencing capacities after >30–50 generations (QA5 is at 100% at each generation tested), which appeared even more stable than those reported in de Vanssay *et al.* (2012). One line (BB5) exhibited a period of strong silencing capacity followed by a reduction in silencing and one line showed partial silencing capacities (BB6). For the *BX2** lines activated by piRNAs homologous to the *P-white* transgene (called *RS3BX2**), one line (QA8) established a long-term strong level of repression, while the other line (BB7-4) showed diminishing repression over time. For the *BX2** lines activated by piRNAs homologous to both *P-1152* and *RS3* transgenes (called *P-1152 + RS3BX2**), the situation resembled that of lines activated by *P-1152* cytoplasm, with one line showing complete repression capacity (QA1) and other lines showing intermediate repression levels. An additive effect was not observed since these *P-1152 + RS3BX2** lines did not appear stronger than the *P-1152BX2** lines despite the fact that small RNAs covering most of the sequence were maternally transmitted in *G*₀. We conclude that paramutation by partially homologous transgenes is globally less efficient than para-

mutation by fully homologous transgenes. However, piRNAs produced by partially homologous transgenes can induce stable and strong paramutation we call “partially homologous paramutations.”

For each type of *G*₀ maternal inheritance, lines showing strong silencing capacities were chosen and ovarian small RNA production was analyzed in *G*₃, *G*₅ and *G*₁₀ by deep sequencing (Figure 3C and Figure S5). Figure 3C shows that, as early as generation 3, the three types of *G*₀ maternal inheritance resulted in strong production of small RNAs covering the entire *BX2 P{lacW}* transgene length. Thus, partially homologous paramutations are associated with rapid and stable spreading of piRNA production by nonhomologous sequences of the targets, similarly to what was found for target transgenes (Muerdter *et al.* 2011; Olovnikov *et al.* 2013; Han *et al.* 2015; Mohn *et al.* 2015) or transposable elements (Shpiz *et al.* 2014).

Further, the paramutagenicity of *P-1152BX2**, *RS3BX2**, and *P-1152+RS3BX2** females was tested. Figure 4 shows that maternal transmission of cytoplasm produced by these females, combined to paternal transmission of a *BX2* naive locus, resulted in female progeny showing strong *trans*-silencing properties. These capacities are transmitted over generations, as tested until *G*₃₆. Thus, paramutation mediated via maternal inheritance of piRNAs homologous to a subregion of the targeted locus can be strong and stable and can generate strongly paramutagenic loci.

Maternally transmitted small RNAs can affect an adult somatic phenotype

piRNAs are maternally transmitted not only to the posterior pole of *Drosophila* embryos but also to more anterior regions where they can be implicated in gene regulation. Indeed, piRNAs homologous to transposable elements (*roo*, 412) have been implicated in *nanos* mRNA decay at the anterior part of the embryo by a mechanism involving mRNA deadenylation (Rouget *et al.* 2010). Furthermore, a high number of maternal RNAs were shown to undergo piRNA-dependent destabilization in the soma during maternal-to-zygotic transition in embryos (Barckmann *et al.* 2015). In addition, piRNAs can have an effect in somatic cells of the gonads. Indeed, in *D. simulans*, expression of the *tirant* transposable element in ovarian follicle cells was shown to be affected by *tirant* homologous piRNA maternal transmission (Akkouche *et al.* 2013). Finally, maternal depletion of Piwi in *D. melanogaster* impacted *white* variegation in adult eyes (Gu and Elgin 2013). Therefore, piRNA maternal deposition can impact not only the germline of progeny but also somatic tissues. During the course of experiments performed to study paramutation in the germline using the *T-1* line as paramutagenic line (de Vanssay *et al.* 2012), we detected impact of *P{lacW}* homologous small RNA maternal inheritance on somatic tissues, which we further investigated.

The *T-1* paramutagenic line, which carries seven *P{lacW}* copies inserted in tandem (Dorer and Henikoff 1994, 1997), was shown to produce abundant ovarian piRNAs including

piRNAs homologous to the *white* transformation marker (de Vanssay *et al.* 2012). The *white* gene is involved in eye pigment transport and its loss of function results in white eyes instead of red ones (Figure 5B). The *T-1* cluster shows strong variegation for the *white* marker, i.e., stochastic ON-OFF *white* repression from cell to cell (RIGS) (Dorer and Henikoff 1994, 1997). RIGS is associated with heterochromatin formation on tandem repeats, which includes *de novo* binding of HP1 at the locus (Fanti *et al.* 2003). In the *T-1* line, RIGS-induced variegation was enhanced by X-ray-induced chromosomal rearrangements. We first investigated the potential somatic effects of *P{lacW}* piRNA maternal inheritance by performing reciprocal crosses between *T-1/Cy* and *white* mutant individuals from different genetic backgrounds and comparing female progeny eyes from these crosses (eye images and pigment amounts). Figure 5A shows the very clear difference observed for all the three backgrounds tested (*w*¹¹¹⁸, *w*¹¹¹⁸ *Cy/Xa*, and *yw*^c). Females having maternally inherited the *T-1* cluster presented stronger *white* repression than females having paternally inherited the cluster. Note that in each case the two types of females have strictly the same genotype. The differences were confirmed by analysis of pigment levels (Figure 5A).

We further tested the effect of *P{lacW}* homologous piRNA maternal inheritance on the eye phenotype in flies carrying another *P{lacW}* cluster called *DX1*. This cluster has six tandem *P{lacW}* copies, one of which is in opposite orientation to the others (Dorer and Henikoff 1994, 1997). This cluster shows a low level of variegation at 25° but variegation is stronger at higher temperatures presenting almost a fully white phenotype at 29° (data not shown). *DX1* has no *trans*-silencing capacities (Ronsseray *et al.* 2001; de Vanssay *et al.* 2012). The two reciprocal crosses were first performed between *DX1* and *w*¹¹¹⁸; *Cy/Xa* individuals to test for the existence of a maternal effect of *DX1* inheritance on its eye phenotype and no significant difference was observed between the two types of progeny for *white* expression (Figure 5B). Further *DX1* individuals were crossed with *T-1* individuals and female progeny having inherited *DX1*, but not *T-1*, were analyzed. Figure 5B shows that *DX1* females, having maternally inherited *T-1* cytoplasm, i.e., containing transgenic small RNAs, presented eyes with a stronger level of *white* repression than *DX1* females having inherited *DX1* cytoplasm. Thus, *T-1* maternal cytoplasm induced *DX1* repression in adult heads. Moreover, females produced by crossing *DX1* females and *T-1* males showed significantly stronger *white* repression than those produced by crossing *DX1* females and *w*¹¹¹⁸; *Cy/Xa* males. This likely resulted from genetic background effects since no paternal inheritance from *T-1* *P{lacW}* piRNAs was previously detected (de Vanssay *et al.* 2012). In contrast, the difference observed between the progeny of reciprocal crosses of *DX1* and *T-1* individuals can be attributed to piRNA maternal transmission since these two types of female progeny have the same genotype. Consequently, a repressive effect of a *T-1* cytoplasm, which carries a high level of *P{lacW}* piRNAs, can be observed in

progeny adult eyes with both the *T-1* and *DX1* *P{lacW}* clusters as a read-out. To understand this small RNA somatic impact in progeny, it will be necessary to test it in zygotes, which are mutants for small RNA silencing pathways or chromatin-based repression pathways.

Discussion

The present results establish the paramutated *BX2** locus as a typical piRNA dual-strand cluster whose *de novo* small RNA production involves nuclear and cytoplasmic actors of the piRNA pathway. Among the small RNAs produced by the *BX2** locus, the 23- to 28-nt piRNAs, but not the 21-nt RNAs, are required for paramutation. Finally, a stable paramutation can be generated between partially homologous loci that are not allelic.

Implication of the primary piRNA pathway in *BX2** capacities

Mutant analysis shows thus that *aubergine*, *rhino*, *cutoff*, and *zucchini* are required for a paramutated locus to exhibit piRNA production capacities and *trans*-silencing. Two main molecular models, one nuclear, one cytoplasmic, have been proposed for the paramutation process (de Vanssay *et al.* 2013; Le Thomas *et al.* 2014). Considering a cytoplasmic paramutation model, it can be proposed that a *trans*-generational nuage-to-nuage transfer of information can be the only factor that explains the acquisition of silencing capacities by the *BX2** paramutated locus. It was previously shown by qRT-PCR that the *BX2* naive (nonparamutated) locus is transcribed in both orientations (de Vanssay *et al.* 2012). Thus, during paramutation, piRNAs loaded on Aub or Ago3 in the nuage of *T-1* or *BX2** female nurse cells could be transferred to the nuage of zygotic primordial germ cells (PGCs) having paternally inherited a *BX2* naive locus, resulting in *de novo* targeting of *BX2* *P{lacW}* transcripts by the ping-pong machinery and accounting by itself for the acquisition of silencing capacities by the *BX2** paramutated locus. In a nuclear paramutation model, maternally transmitted piRNAs, loaded on Piwi, could enter into PGC nuclei and activate *de novo* targeting of *BX2* naive transcript by the primary piRNA pathway via promoting the binding of the Rhino-Deadlock-Cutoff complex on the *P{lacW}* cluster. This model is suggested by the fact that deposition of Rhino at several piRNA producing loci depends on Piwi (Mohn *et al.* 2014). It is also strongly supported by ChIP analysis, which showed that a *BX2** cluster, as tested in G₁ ovaries of the paramutation experiment, exhibited an increase in H3K9 methylation levels when compared to a naive *BX2* locus (Le Thomas *et al.* 2014). H3K9me3 residues can indeed be targeted by the chromodomain of Rhino, resulting in RDC complex stabilization on the cluster. The mutant analysis reported here provides the functional assay that shows that *rhino*, *cutoff*, and *zucchini* are necessary for *BX2** *trans*-silencing and piRNA production capacities (Figure 1). Therefore, paramutation-linked repressive properties do not solely depend on cytoplasmic ping-pong

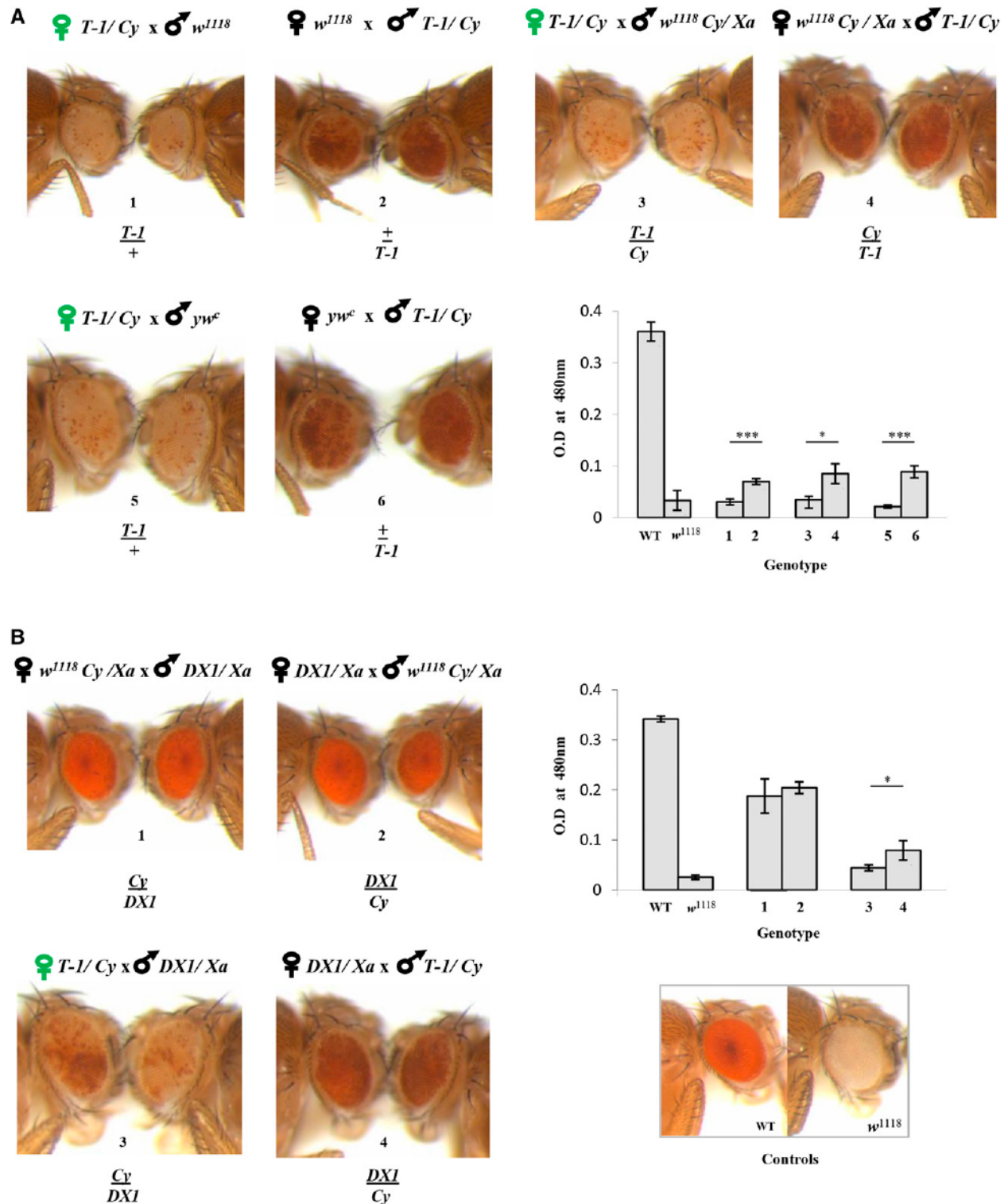


Figure 5 Somatic impact of maternally transmitted small RNAs. The effect of *P{lacW}* small RNA maternal inheritance on *P{lacW}* cluster expression in adult eyes was tested using two different transgene clusters, *T-1* (A) and *DX1* (B), as a read-out. *T-1*, but not *DX1*, has *trans*-silencing capacities, correlated with the presence of abundant ovarian piRNAs. The G_0 cross is indicated above the picture and was performed at 25°. Green female symbol indicates cytoplasmic inheritance containing *P{lacW}* piRNAs. The genotype (chromosome 2) of the G_1 females analyzed is given below the picture. *Cy* and *Xa* are chromosomes devoid of transgenes and carrying dominant markers. The maternally inherited chromosome is written above the bar. (A and B) Pigment levels (absorbance at 480 nm) are shown beside the pictures. Numbers in abscissa indicate the genotype analyzed (reported directly on the head picture). For pigment level dosages, Canton^v and *w¹¹¹⁸* female heads served as WT and *w¹¹¹⁸* female heads served as controls, respectively. Control pictures are boxed in B. Bars indicate the standard error. Comparisons were performed using the Student's *t*-test: **P* < 0.05; ***P* < 0.01; ****P* < 0.001. Maternal inheritance of *P{lacW}* small RNAs produced by a *T-1* female results in increased silencing of both *T-1* and *DX1* expression in the eye.

activation resulting from nuage-to-nuage *trans*-generational transfer of information. These properties depend also on the primary piRNA pathway, which is linked to the binding of the RDC complex on the locus (Mohn *et al.* 2014; Zhang *et al.* 2014). Analysis focused on the *T-1* paramutagenic locus, which produces a high level of ovarian *P{lacW}* piRNAs, strongly reinforces this model (Le Thomas *et al.* 2014). Indeed, ChIP analysis performed on ovaries of *G*₁ females produced by the two reciprocal crosses between *T-1* and wild-type (devoid of transgene) individuals showed that maternal inheritance of the *T-1* cytoplasm is associated with an increase of Rhino and Cutoff binding on the *T-1 P{lacW}* cluster in progeny ovaries (Le Thomas *et al.* 2014). Paramutation associated repressive properties also depend on the Zuc and on Aub activities (Figure 1), which take place in the nuage. In conclusion, chromatin analyses (Le Thomas *et al.* 2014) and functional analyses performed with mutants (Figure 1) complement each other to show that *trans*-generational induction of piRNA-mediated silencing by maternally inherited piRNAs involves nuclear and cytoplasmic steps of the piRNA pathway, including those involved in the atypical transcription of the piRNA loci. Thus, during paramutation, the paramutated locus is modified and remains so over generations.

BX2* transgenic 21-nt RNAs are actual siRNAs whose production also depends on the piRNA machinery

Deep sequencing of ovarian small RNAs associated with both telomeric *P*-transgene insertions and to the *T-1* paramutagenic or *BX2** paramutated clusters showed abundant 21-nt transgene homologous RNAs, in addition to 23- to 28-nt piRNAs (de Vanssay *et al.* 2012). Both sense and antisense 21-nt RNAs were found that contained no ping-pong signatures. These 21-nt RNAs were proposed to be *Dicer-2* dependent. Here, ovarian small RNAs of *BX2**, *Dcr-2*^{-/-} females have been analyzed by deep sequencing. As expected, *BX2** 21-nt RNAs were almost absent in these *Dcr-2* mutant females showing that they are genuine siRNAs (Figure 1). Strikingly, all the piRNA pathway mutations tested also resulted in the loss of the transgene-derived 21-nt RNAs (Figure 1). This suggests that *P{lacW}* siRNAs could be generated downstream of the piRNA pathway. It was shown further that production of these transgene-derived siRNAs by a female requires piRNA, but not siRNA, maternal inheritance, confirming that their appearance is downstream of piRNAs (Figure 2). The 21-nt small RNA production appears to be a common feature for a number of piRNA-producing loci (Czech *et al.* 2008; Malone *et al.* 2009; Pane *et al.* 2011). *cuff* mutant analysis was shown to strongly affect 21-nt RNA production by a number of double-strand piRNA-producing loci, including the 42AB locus (Pane *et al.* 2011; Le Thomas *et al.* 2014). What mechanism could explain the 21-nt RNA dependence on the piRNA pathway? *P{lacW}* homologous siRNA biogenesis depends on all the piRNA actors tested, including for example Rhino, which is necessary for production of piRNA precursors; Zucchini, which is responsible for slicing

steps in the cytoplasm; and Aubergine, which is an actor of the ping-pong amplification of the pool of piRNA. One possibility is that *P{lacW}* siRNA production results from the fact that both sense and antisense RNAs, produced by this dual-strand cluster, are concentrated locally due to the piRNA machinery, for example in the nuage, leading to occurrence of double-strand RNAs, which can further be processed by *Dcr-2*. The piRNA mutants tested could alter the production of the piRNAs associated with Piwi, which are produced downstream of the ping-pong amplification step (Han *et al.* 2015; Mohn *et al.* 2015), and consequently reduce a positive feedback in the nucleus of these piRNAs (Mohn *et al.* 2014). This would reduce deposition of Rhino at several piRNA, producing loci (Mohn *et al.* 2014) and thus affect dual-strand transcription of the *BX2* locus, abolish the sense and antisense RNA concentration step, and siRNA production as well. However, a low level of *P{lacW}* 21-nt RNA is still detectable in *Dcr-2* homozygous mutant females (Figure 1F), which present a ping-pong signature (Figure S3J). If we analyze small RNAs produced by the 42AB locus, which is a typical dual-strand cluster piRNA locus, piRNAs are reduced or disappear in all piRNA pathway mutant backgrounds tested (*aub*^{-/-}, *rhi*^{-/-}, *cuff*^{-/-}, *zuc*^{-/-}, Figure S1, D, H, J, and L). The 21-nt RNAs with a ping-pong signature also disappear in these mutant backgrounds for the 42AB locus (Figure S3, H, P, T, and X). An intriguing variation is observed in the *zuc*^{-/-} background (Figure S1L), which showed reduction in piRNA levels and a dramatic increase in 21-nt RNA levels for the 42AB locus. This suggests that in the absence of Zuc, precursor transcripts coming from double-stranded piRNA clusters can be processed into 21-nt RNA instead of piRNAs. This could be achieved by *Dcr-2*, since this enzyme is responsible for the presence of the majority of the 21-nt RNAs coming from piRNA clusters (see above) and these 21-nt RNAs show no ping-pong signature (Figure S3X). Thus, for a given locus, two 21-nt RNA populations coexist in the *Drosophila* germline: a high level of *Dcr-2*-dependent siRNAs showing no ping-pong signature whose production could be in competition with that of piRNAs and a low level of *Dcr-2* independent 21-nt piRNAs showing a ping-pong signature.

Spreading of the piRNA production capacity within a piRNA-producing cluster can be completed within three generations

piRNA clusters behave as genomic traps in which new insertions of TEs lead to *de novo* production of piRNA homologous to the new insertions. The kinetics of appearance of this new piRNA production is not clear, but some functional evidence has been provided from previous studies concerning the repression capacity inheritance of the *I*-factor and of the *P*-element (Kidwell *et al.* 1977; Bucheton 1978; Engels 1979). In both systems, it appears that establishment of repressive capacities, following paternal inheritance of *I*- or *P*-elements carrying genomes in a cytoplasm devoid of *I*- or *P*-homologous piRNAs, is progressive and require several generations to be completed (Engels 1979, 1989; Bregliano *et al.* 1980;

Finnegan 1989; Rio 2002). For example, experiments were performed with a line [*Lk-P(1A)*] carrying only two regulatory *P*-elements inserted in a telomeric strong piRNA-producing locus (TAS) (Coen *et al.* 1994; Ronsseray *et al.* 1996). It was shown that paternal transmission of the telomeric *P*-element locus in a cytoplasm devoid of *P*-homologous piRNAs, followed by maternal inheritance of the telomeric *P*-elements in successive generations, requires four generations to progressively establish complete *P*-repression (Coen *et al.* 1994). Thus, during these generations, a progressive spreading must occur within the TAS locus for piRNA production capacity, which finally results in strong production of piRNAs homologous, not only to the TAS, but also to the *P*-element. A spreading process for piRNA production was also observed for the *flamenco* piRNA-producing locus, which functions in the somatic follicle cells that surround germline cysts (Muerdter *et al.* 2011). Further, analysis of previously established transgenic lines that carry an *I*-fragment within a *P*-transgene have shown that piRNA production capacity can spread within transgenes (Olovnikov *et al.* 2013). The capacity to produce piRNAs invaded the whole *I*-derived transgene and reached in some cases flanking sequences. A similar spreading process into flanking sequences was also found for natural transposon insertions (Shpiz *et al.* 2014). It is possible, however, that spreading processes inside or outside a transgene or a TE involve different mechanisms. However, in these last experiments, the capacity of targeted sequences to maintain piRNA production on their own, in absence of the endogenous homologous piRNA locus that stimulated them, was not tested. In this article, we show that *cis*-spreading of piRNA production on *BX2** is maintained over generations in absence of the original input, the telomeric transgenic locus. In addition, this spreading is completed within a few generations.

What mechanism could be responsible for the spreading process? Very recent studies have shown that piRNAs associated with Piwi can be produced downstream of piRNA generation by ping-pong amplification, which can increase piRNA sequence diversity (Han *et al.* 2015; Mohn *et al.* 2015). Indeed, piRNAs associated with Ago3 can target complementary RNA and induce production of phased piRNAs (27 nt) corresponding to the slicing of the 3' tail of the cleaved target RNAs. These phased "trailer" piRNA, once loaded on Piwi, can potentially return to the nucleus and increase both transcriptional repression of homologous targeted TEs or stimulate binding of Rhino on the piRNA-producing locus. One consequence of such an additional mechanism would be the adjustment of the force of repression to the level of transcription of active copies. Another would be the formation of novel piRNAs, which are *cis* to the sequence targeted initially by piRNAs. This phenomenon would explain why a *P-lacZ-white* transcript targeted by *P-lacZ* homologous piRNAs is able to produce piRNAs homologous to *white*. However, some observations remain to be explained. First, according to this model, when telomeric *P*-insertions are paternally introduced in a lineage and thereupon are maternally transmitted, establishment

of repression would be expected to require only one generation to be completed but in fact it was shown to be progressive over generations (Coen *et al.* 1994). It is possible that nuclear reinforcement mediated by trailer piRNAs loaded on Piwi on the piRNA-producing locus is necessary, and this reinforcement involves maternal transmission to be efficient and thus requires more than one generation to be completed. Second, we performed an additional paramutation experiment in which a paternally inherited *BX2* locus was introduced in a maternal cytoplasm produced by *Lk-P(1A)* females. In that experiment, no paramutation of the paternally transmitted *BX2* cluster was observed (Figure S6) despite maternal inheritance of *P*-element homologous piRNAs corresponding to 820 bp of homology of the 10,691 bp of the *P{lacW}* transgene. It remains possible, however, that production of such trailer piRNAs is limited by the propagation mechanism itself, making it possible to fill a gap of 4 kb (Figure 3) but not a gap of almost 10 kb (Figure S6).

Role of paramutation in TE repression establishment and stabilization in natural populations

The *Drosophila* genome contains >140 piRNA-producing loci, most of them composed of truncated versions of TEs (Brennecke *et al.* 2007). How have these loci been activated in natural populations to produce piRNAs? Paramutation may have played a key role in such a process. Once activated in a single fly by an unknown factor, which could be environmental (Schott *et al.* 2014), a first piRNA-producing locus may have progressively replaced the inactive allele within the population by a paramutation process similar to paramutation of a *BX2* naive locus by a *BX2** locus (de Vanssay *et al.* 2012). Further, the first locus activated may have activated other TE-accumulating loci, despite showing only partial homology and being located on a different chromosome, by a paramutation process similar to that of *BX2* by *P-1152* or *RS3* (Figure 3). These secondarily activated sites could maintain TE repression capacities in individuals from the population that did not inherit the first active piRNA locus if this first locus is not fixed in the population. With time, epigenetic communication between TE-accumulating loci may activate all "genetically" competent loci (sufficient number of tandem repeats, bidirectional transcription, etc.). Moreover, these partially homologous paramutations could also have played a role during migration processes between natural populations. Indeed, a fly bearing active loci arriving from one population into another one could activate loci present on the chromosomes of the recipient population, despite the existence of potential partial divergences between TE-accumulating loci in these different populations. Therefore, it is possible to propose that paramutation may play an important role in epigenome shaping of the *Drosophila* germline, especially because it can occur even when interacting loci partially diverge in their sequence.

Acknowledgements

We thank Doug Dorer, Steve Henikoff, Attilio Pane, Trudi Schüpbach, Paul Macdonald, and the Bloomington Stock

Center for providing stocks and FlyBase.org for providing databases. We thank Ritha Zamy for technical assistance, Neel Randsholt and Anne-Marie Pret for critical reading of the manuscript, and an anonymous referee for useful suggestions. This work was supported by fellowships from the Ministère de l'Enseignement Supérieur et de la Recherche to C.H. and A.A.-L. and by grants from the Association de la Recherche contre le Cancer (Fondation ARC, SFI20121205921, SFI20131200470), the Fondation pour la Recherche Médicale (FRM DEP20131128532), the Association Nationale de la Recherche (ANR, project "plastisipi") to S.R. and C.A., and the University Pierre et Marie Curie (Emergence EME1223) to L.T. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript. The authors declare that no competing interest exists.

Literature Cited

- Akkouche, A., T. Grentzinger, M. Fablet, C. Armenise, N. Burlet *et al.*, 2013 Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells. *EMBO Rep.* 14: 458–464.
- Antoniewski, C., 2014 Computing siRNA and piRNA overlap signatures. *Methods Mol. Biol.* 1173: 135–146.
- Barckmann, B., S. Pierson, J. Dufourt, C. Papin, C. Armenise *et al.*, 2015 Aubergine iCLIP reveals piRNA-dependent decay of mRNAs involved in germ cell development in the early embryo. *Cell Reports* 12: 1205–1216.
- Batista, P. J., J. G. Ruby, J. M. Claycomb, R. Chiang, N. Fahlgren *et al.*, 2008 PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol. Cell* 31: 67–78.
- Biemont, C., S. Ronsseray, D. Anxolabehere, H. Izaabel, and C. Gautier, 1990 Localization of P elements, copy number regulation, and cytotype determination in *Drosophila melanogaster*. *Genet. Res.* 56: 3–14.
- Bregliano, J. C., G. Picard, A. Bucheton, A. Pelisson, J. M. Lavigne *et al.*, 1980 Hybrid dysgenesis in *Drosophila melanogaster*. *Science* 207: 606–611.
- Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis *et al.*, 2007 Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128: 1089–1103.
- Brennecke, J., C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark *et al.*, 2008 An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322: 1387–1392.
- Brink, R. A., 1956 A genetic change associated with the R locus in maize which is directed and potentially reversible. *Genetics* 41: 872–889.
- Bucheton, A., 1978 Non-Mendelian female sterility in *Drosophila melanogaster*: influence of ageing and thermic treatments. I. Evidence for a partly inheritable effect of these two factors. *Heredity* 41: 357–369.
- Chandler, V. L., 2007 Paramutation: from maize to mice. *Cell* 128: 641–645.
- Coe, E. H., 1959 A regular and continuing conversion-type phenomenon at the B locus in maize. *Proc. Natl. Acad. Sci. USA* 45: 828–832.
- Coen, D., B. Lemaitre, M. Delattre, H. Quesneville, S. Ronsseray *et al.*, 1994 *Drosophila* P element: transposition, regulation and evolution. *Genetica* 93: 61–78.
- Czech, B., C. D. Malone, R. Zhou, A. Stark, C. Schlingehayde *et al.*, 2008 An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453: 798–802.
- de Vanssay, A., A.-L. Bouge, A. Boivin, C. Hermant, L. Teyssset *et al.*, 2012 Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature* 490: 112–115.
- de Vanssay, A., A. L. Bouge, A. Boivin, C. Hermant, L. Teyssset *et al.*, 2013 piRNAs and epigenetic conversion in *Drosophila*. *Fly (Austin)* 7: 237–41.
- Dorer, D. R., and S. Henikoff, 1994 Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell* 77: 993–1002.
- Dorer, D. R., and S. Henikoff, 1997 Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. *Genetics* 147: 1181–1190.
- Dufourt, J., C. Dennis, A. Boivin, N. Gueguen, E. Theron *et al.*, 2014 Spatio-temporal requirements for transposable element piRNA-mediated silencing during *Drosophila* oogenesis. *Nucleic Acids Res.* 42: 2512–2524.
- Engels, W. R., 1979 Hybrid dysgenesis in *Drosophila melanogaster*: rules of inheritance of female sterility. *Genet. Res.* 33: 219–236.
- Engels, W. R., 1989 *P Elements in Drosophila*, American Society for Microbiology, Washington, DC.
- Fanti, L., M. Berloco, L. Piacentini, and S. Pimpinelli, 2003 Chromosomal distribution of heterochromatin protein 1 (HP1) in *Drosophila*: a cytological map of euchromatic HP1 binding sites. *Genetica* 117: 135–147.
- Finnegan, D. J., 1989 The I factor and I-R hybrid dysgenesis in *Drosophila melanogaster*. *Mobile DNA Chp.* 18: 503–518.
- Grentzinger, T., C. Armenise, C. Brun, B. Mugat, V. Serrano *et al.*, 2012 piRNA-mediated transgenerational inheritance of an acquired trait. *Genome Res.* 22: 1877–1888.
- Gu, T., and S. C. Elgin, 2013 Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in *Drosophila*. *PLoS Genet.* 9: e1003780.
- Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura *et al.*, 2007 A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315: 1587–1590.
- Han, B. W., W. Wang, C. Li, Z. Weng, and P. D. Zamore, 2015 Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* 348: 817–821.
- Ipsaro, J. J., A. D. Haase, S. R. Knott, L. Joshua-Tor, and G. J. Hannon, 2012 The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* 491: 279–283.
- Iwasaki, Y. W., M. C. Siomi, and H. Siomi, 2015 PIWI-interacting RNA: its biogenesis and functions. *Annu. Rev. Biochem.* 84: 405–433.
- Josse, T., L. Teyssset, A. L. Todeschini, C. M. Sidor, D. Anxolabehere *et al.*, 2007 Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS Genet.* 3: 1633–1643.
- Josse, T., C. Maurel-Zaffran, A. de Vanssay, L. Teyssset, A. L. Todeschini *et al.*, 2008 Telomeric trans-silencing in *Drosophila melanogaster*: tissue specificity, development and functional interactions between non-homologous telomeres. *PLoS One* 3: e3249.
- Karpen, G. H., and A. C. Spradling, 1992 Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics* 132: 737–753.
- Kidwell, M. G., J. F. Kidwell, and J. A. Sved, 1977 Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility, and male recombination. *Genetics* 86: 813–833.
- Klattenhoff, C., H. Xi, C. Li, S. Lee, J. Xu *et al.*, 2009 The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* 138: 1137–1149.
- Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10: R25.
- Le Thomas, A., A. K. Rogers, A. Webster, G. K. Marinov, S. E. Liao *et al.*, 2013 Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* 27: 390–399.

- Le Thomas, A., E. Stuwe, S. Li, J. Du, G. Marinov *et al.*, 2014 Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev.* 28: 1667–1680.
- Lee, Y. S., K. Nakahara, J. W. Pham, K. Kim, Z. He *et al.*, 2004 Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 117: 69–81.
- Lemaitre, B., S. Ronsseray, and D. Coen, 1993 Maternal repression of the P element promoter in the germline of *Drosophila melanogaster*: a model for the P cytotype. *Genetics* 135: 149–160.
- Malone, C. D., J. Brennecke, M. Dus, A. Stark, W. R. McCombie *et al.*, 2009 Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137: 522–535.
- Mohn, F., G. Sienski, D. Handler, and J. Brennecke, 2014 The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* 157: 1364–1379.
- Mohn, F., D. Handler, and J. Brennecke, 2015 Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* 348: 812–817.
- Muerdter, F., I. Olovnikov, A. Molaro, N. V. Rozhkov, B. Czech *et al.*, 2011 Production of artificial piRNAs in flies and mice. *RNA* 18: 42–52.
- Nishimasu, H., H. Ishizu, K. Saito, S. Fukuhara, M. K. Kamatani *et al.*, 2012 Structure and function of Zucchini endonuclease in piRNA biogenesis. *Nature* 491: 284–287.
- O’Kane, C. J., and W. J. Gehring, 1987 Detection in situ of genomic regulatory elements in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 84: 9123–9127.
- Olovnikov, I., S. Ryazansky, S. Shpiz, S. Lavrov, Y. Abramov *et al.*, 2013 De novo piRNA cluster formation in the *Drosophila* germ line triggered by transgenes containing a transcribed transposon fragment. *Nucleic Acids Res.* 41: 5757–5768.
- Pane, A., P. Jiang, D. Y. Zhao, M. Singh, and T. Schüpbach, 2011 The Cutoff protein regulates piRNA cluster expression and piRNA production in the *Drosophila* germline. *EMBO J.* 30: 4601–4615.
- Pöyhönen, M., A. de Vanssay, V. Delmarre, C. Hermant, A. L. Todeschini *et al.*, 2012 Homology-dependent silencing by an exogenous sequence in the *Drosophila* germline. *G3 (Bethesda)* 2: 331–338.
- Rio, D. C., 2002, pp. 484–518 in *P Transposable Elements in Drosophila melanogaster. Mobile DNA II*, edited by N. L. Craig, *et al.*, American Society for Microbiology, Washington, DC.
- Roche, S. E., and D. C. Rio, 1998 Trans-silencing by P elements inserted in subtelomeric heterochromatin involves the *Drosophila* Polycomb group gene, Enhancer of zeste. *Genetics* 149: 1839–1855.
- Ronsseray, S., M. Lehmann, and D. Anxolabehere, 1991 The maternally inherited regulation of P elements in *Drosophila melanogaster* can be elicited by two P copies at cytological site 1A on the X chromosome. *Genetics* 129: 501–512.
- Ronsseray, S., M. Lehmann, D. Nouaud, and D. Anxolabehere, 1996 The regulatory properties of autonomous subtelomeric P elements are sensitive to a Suppressor of variegation in *Drosophila melanogaster*. *Genetics* 143: 1663–1674.
- Ronsseray, S., A. Boivin, and D. Anxolabehere, 2001 P-element repression in *Drosophila melanogaster* by variegating clusters of P-lacZ-white transgenes. *Genetics* 159: 1631–1642.
- Ronsseray, S., T. Josse, A. Boivin, and D. Anxolabehere, 2003 Telomeric transgenes and trans-silencing in *Drosophila*. *Genetica* 117: 327–335.
- Rouget, C., C. Papin, A. Boureux, A. C. Meunier, B. Franco *et al.*, 2010 Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. *Nature* 467: 1128–1132.
- Rozhkov, N. V., M. Hammell, and G. J. Hannon, 2013 Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes Dev.* 27: 400–412.
- Ruby, J. G., C. Jan, C. Player, M. J. Axtell, W. Lee *et al.*, 2006 Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127: 1193–1207.
- Sapetschnig, A., P. Sarkies, N. J. Lehrbach, and E. A. Miska, 2015 Tertiary siRNAs mediate paramutation in *C. elegans*. *PLoS Genet.* 11: e1005078.
- Sato, K., and M. C. Siomi, 2013 Piwi-interacting RNAs: biological functions and biogenesis. *Essays Biochem.* 54: 39–52.
- Schott, D., I. Yanai, and C. P. Hunter, 2014 Natural RNA interference directs a heritable response to the environment. *Sci. Rep.* 4: 7387.
- Schupbach, T., and E. Wieschaus, 1991 Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* 129: 1119–1136.
- Shirayama, M., M. Seth, H. C. Lee, W. Gu, T. Ishidate *et al.*, 2012 piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell* 150: 65–77.
- Shpiz, S., S. Ryazansky, I. Olovnikov, Y. Abramov, and A. Kalmykova, 2014 Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the *Drosophila* germline. *PLoS Genet.* 10: e1004138.
- Sienski, G., D. Donertas, and J. Brennecke, 2012 Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151: 964–980.
- Siomi, H., and M. C. Siomi, 2015 RNA. Phased piRNAs tackle transposons. *Science* 348: 756–757.
- Todeschini, A. L., L. Teyssset, V. Delmarre, and S. Ronsseray, 2010 The epigenetic trans-silencing effect in *Drosophila* involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1. *PLoS One* 5: e11032.
- Voigt, F., M. Reuter, A. Kasaruho, E. C. Schulz, R. S. Pillai *et al.*, 2012 Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc. *RNA* 18: 2128–2134.
- Volpe, A. M., H. Horowitz, C. M. Grafer, S. M. Jackson, and C. A. Berg, 2001 *Drosophila* rhino encodes a female-specific chromodomain protein that affects chromosome structure and egg polarity. *Genetics* 159: 1117–1134.
- Wang, S. H., and S. C. Elgin, 2011 *Drosophila* Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc. Natl. Acad. Sci. USA* 108: 21164–21169.
- Weick, E. M., and E. A. Miska, 2014 piRNAs: from biogenesis to function. *Development* 141: 3458–3471.
- Weiler, K. S., and B. T. Wakimoto, 2002 Suppression of heterochromatic gene variegation can be used to distinguish and characterize E(var) genes potentially important for chromosome structure in *Drosophila melanogaster*. *Mol. Genet. Genomics* 266: 922–932.
- Zhang, Z., J. Wang, N. Schultz, F. Zhang, S. S. Parhad *et al.*, 2014 The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* 157: 1353–1363.

Communicating editor: J. Birchler

GENETICS

Supporting Information

www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.180307/-/DC1

Paramutation in *Drosophila* Requires Both Nuclear and Cytoplasmic Actors of the piRNA Pathway and Induces *Cis*-spreading of piRNA Production

Catherine Hermant, Antoine Boivin, Laure Teyssset, Valérie Delmarre, Amna Asif-Laidin, Marius van den Beek, Christophe Antoniewski, and Stéphane Ronsseray

Paramutation in *Drosophila* requires both nuclear and cytoplasmic actors of the piRNA pathway and induces *cis*-spreading of piRNA production.

Catherine HERMANT et al., Genetics

Corresponding author: stephane.ronsseray@upmc.fr

SUPPORTING INFORMATION

(Table S1, Figures S1-S6)

The nature of genetic variation for complex traits revealed by GWAS and regional heritability mapping analyses, pp. 1601–1613

Armando Caballero, Albert Tenesa, and Peter D. Keightley

Caballero *et al.* used simulations to show that, contrary to previous results, common variants of large effect are responsible for most of the genetic variation for quantitative traits (except when the trait is fitness itself). They also show that knowledge of full sequence data and multi-SNP methods are unlikely to substantially reduce missing heritability, because additional QTLs revealed by more powerful methods only modestly increase the proportion of the heritability explained.

Paramutation in *Drosophila* requires both nuclear and cytoplasmic actors of the piRNA pathway and induces *cis*-spreading of piRNA production, pp. 1381–1396

Catherine Hermant, Antoine Boivin, Laure Teyssset, Valérie Delmarre, Amna Asif-Laidin, Marius van den Beek, Christophe Antoniewski, and Stéphane Ronsseray

Mobile DNA is repressed in animal gonads by non-coding small RNAs called Piwi-Interacting RNAs (piRNAs). In *Drosophila* the capacity of a genomic locus to produce piRNAs can be transferred from a producing locus to the allelic previously non-producing homologous locus, resulting in an epigenetic conversion process termed paramutation. These authors show that this paramutation event involves nuclear and cytoplasmic components of the piRNA machinery. In addition, since paramutagenic conversion is also shown to occur between partially homologous non-allelic loci, it likely plays an important role in shaping of the epigenome in the wild.

A gene regulatory program in human breast cancer, pp. 1341–1348

Renhua Li, John Campos, and Joji Iida

Li *et al.* identified master regulator genes that shape different tumor subtypes in human breast cancer. Using a machine learning algorithm, they identified 16 such genes. On the basis of gene expression patterns across three large cohorts of human breast cancer populations, the genes can be congruently divided into two groups that regulate cancer related genes in opposite directions. Their studies define a gene regulatory program for tumor progression.

Selective strolls: fixation and extinction in diploids are slower for weakly selected mutations than for neutral ones, pp. 1581–1589

Fabrizio Mafessoni and Michael Lachmann

A common assumption among geneticists is that neutral alleles survive longer in a population than selected variants: negative selection rapidly leads to extinction of deleterious mutations, while advantageous alleles spread faster till fixation. This paper shows that these assumptions are often incorrect. Weakly selected mutations behave in a fashion opposite to strongly selected ones, with slower average fixation and extinction times than neutral alleles. Hence, fixation events due to weak selection will show opposite patterns of statistics used to detect signatures of positive selection.

Roles of nucleoid-associated proteins in stress-induced mutagenic break repair in starving *Escherichia coli*, pp. 1349–1362

Jessica M. Moore, David Magnan, Ana K. Mojica, María Angélica Bravo Núñez, David Bates, Susan M. Rosenberg, and P. J. Hastings

Like eukaryotic chromosomes, bacterial nucleoids have DNA compaction proteins. These nucleoid-associated proteins are global transcriptional regulators. These investigators show that nucleoid-associated proteins modulate mutagenesis. Mutagenic break repair in *Escherichia coli* is regulated by proteins that detect stress and activate proteins that act on DNA to make mutations, potentially accelerating evolution during stress. Moore *et al.* found that nine of fifteen nucleoid-associated proteins are required for mutagenic break repair; one inhibits it. Six of them regulate stress-responses, which couple mutagenesis to stress and create genetic diversity

Cell differentiation and spatial organization in yeast colonies: role of cell-wall integrity pathway, pp. 1427–1438

Sarah Piccirillo, Rita Morales, Melissa G. White, Keston Smith, Tamas Kapros, and Saul M. Honigberg

This paper reports that yeast colonies partition into two distinct cell layers: an underlying layer of permeable “feeder” cells and an overlying layer of cells undergoing sexual reproduction (meiosis). Activation of a MAPK pathway in the feeder cell layer stimulates cell permeability and indirectly activates meiosis in the upper layer. In stressful or suboptimal environments, more cells in the colony adopt a feeder cell fate, and under these conditions feeder cell development and MAPK pathway activity become more important for sexual development in the overlying layer.

Manipulation of karyotype in *Caenorhabditis elegans* reveals multiple inputs driving pairwise chromosome synapsis during meiosis, pp. 1363–1379

Baptiste Roelens, Mara Schvarzstein, and Anne M. Villeneuve

Roelens *et al.* devised an efficient strategy for generating polyploid derivatives of virtually any *Caenorhabditis elegans* strain, then exploited this ability to manipulate ploidy to experimentally challenge the meiotic program. Their data suggest that three separate inputs contribute to stable pairwise associations (synapsis) between homologous chromosomes during meiosis: 1) recombination-independent assessment of homology near special chromosome sites known as pairing centers, 2) a strong driving force favoring pairwise over multi-partner interactions, and 3) recombination-dependent maturation of homologous synapsis.

Associating multivariate quantitative phenotypes with genetic variants in family samples with a novel kernel machine regression method, pp. 1329–1339

Qi Yan, Daniel E. Weeks, Juan C. Celedón, Hemant K. Tiwari, Bingshan Li, Xiaojing Wang, Wan-Yu Lin, Xiang-Yang Lou, Guimin Gao, Wei Chen, and Nianjun Liu

Recent developments in sequencing technology enables genetic variants to be associated with complex diseases. Jointly testing for association between genetic variants and multiple correlated phenotypes may increase the power to detect causal genes in family-based studies, but familial correlation needs to be appropriately handled. Yan *et al.* propose a novel approach that is based on a linear mixed model framework that can be applied to a large range of studies with different types of traits.

This Month in the American Journal of Human Genetics

Multiple hepatic regulatory variants at the GALNT2 high-density lipoprotein cholesterol GWAS locus, Am. J. Hum. Genet. 98(1)

Tamara S. Roman, Amanda F. Marvelle, Marie P. Fogarty, Swarooparani Vadlamudi, Arlene J. Gonzalez, Martin L. Buchkovich, Jeroen R. Huyghe, Christian Fuchsberger, Anne U. Jackson, Ying Wu, Mete Civelek, Aldons J. Lusis, Kyle J. Gaulton, Praveen Sethupathy, Antti J. Kangas, Pasi Soininen, Mika Ala-Korpela, Johanna Kuusisto, Francis S. Collins, Markku Laakso, Michael Boehnke, and Karen L. Mohlke

GWAS has produced an enormous number of SNPs associated with various traits, but the function of many of these signals is unclear. Experimental follow-up is required to determine the functional variants underlying the signal and the consequences of variation at these sites. In this study, Roman *et al.* fine map a region near *GALNT2*, that is associated with blood lipids and cholesterol levels. At least two variants in LD with the tag-SNP appear to alter transcription factor binding and *GALNT2* expression. Together, these results demonstrate that multiple regulatory variants near *GALNT2*, encoding a gene involved in HDL-C metabolism, might explain the GWAS signal and suggest that future work to determine how *GALNT2* influences HDL particle size could be useful in understanding blood lipids in humans.

Library ID	1		2		3		4		5		6		7	
Library	GRH12		GRH49		GRH50		GRH20		GRH19		GRH13		GRH14	
Parental origin	<i>BX2/Cy</i> x <i>Canton'</i>		<i>BX2*/CyRoi</i> x <i>w¹¹¹⁸</i>		<i>BX2*/CyRoi</i> x <i>w¹¹¹⁸</i>		<i>BX2*,aub^{0C42}/Cy</i> x <i>aub^{N11}/Xa</i>		<i>BX2*,aub^{0C42}/Cy</i> x <i>aub^{N11}/Xa</i>		<i>BX2*,Dcr-2^{1116X}/Cy</i> x <i>Dcr-2^{1116X}/Cy</i>		<i>BX2*,Dcr-2^{1116X}/Cy</i> x <i>Dcr-2^{1116X}/Cy</i>	
Female genotype	<i>BX2/+</i>		<i>BX2*_{G83}/+</i>		<i>CyRoi/+</i>		<i>BX2*,aub^{0C42}/Xa</i>		<i>BX2*,aub^{0C42}/aub^{N11}</i>		<i>BX2*,Dcr-2^{1116X}/Cy</i>		<i>BX2*,Dcr-2^{1116X}/Dcr-2^{1116X}</i>	
Dmel_matched	20 140 166	100.00%	15 675 824	100.00%	14 222 359	100.00%	12 774 247	100.00%	11 504 770	100.00%	25 425 639	100.00%	17 615 555	100.00%
miRNAs	6 848 463	34.00%	3 867 491	24.67%	3 207 689	22.55%	1 536 928	12.03%	2 097 790	18.23%	4 570 877	17.98%	3 659 767	20.78%
tRNAs	1 101 672	5.47%	365 007	2.33%	208 814	1.47%	129 576	1.01%	205 847	1.79%	1 044 735	4.11%	960 156	5.45%
ncRNAs	158 102	0.79%	97 362	0.62%	94 152	0.66%	103 080	0.81%	95 869	0.83%	130 168	0.51%	120 312	0.68%
miscRNAs	1 393 137	6.92%	763 262	4.87%	924 378	6.50%	7 952 990	62.26%	6 939 045	60.31%	6 909 892	27.18%	3 156 238	17.92%
Transposons	7 687 715	38.17%	8 021 967	51.17%	7 315 480	51.44%	2 184 449	17.10%	1 426 029	12.40%	9 460 422	37.21%	7 189 726	40.81%
Introns	935 677	4.65%	784 756	5.01%	754 263	5.30%	240 681	1.88%	190 169	1.65%	1 105 726	4.35%	827 244	4.70%
Transcripts	527 498	2.62%	351 601	2.24%	365 592	2.57%	200 782	1.57%	201 707	1.75%	540 473	2.13%	457 607	2.60%
Intergenic	1 307 958	6.49%	1 294 915	8.26%	1 217 106	8.56%	379 149	2.97%	2 88 815	2.51%	1 494 552	5.88%	1 117 847	6.35%
Remaining	179 944	0.89%	129 463	0.83%	134 885	0.95%	46 612	0.36%	59 499	0.52%	168 794	0.66%	126 658	0.72%
Effective depth	1 7645 357		14 547 555		13 089 167		4 691 681		4 359 878		17 471 012		13 499 161	
Normalisation factor	0.2470835		0.2996983		0.3330905		0.9292784		1		0.2495492		0.322974	

Library ID	8		9		10		11		12		13	
Library	GRH60		GRH61		GRH44		GRH43		GRH34		GRH33	
Parental origin	<i>BX2*,rhl²/Cy</i> x <i>rhl^{KG}/Cy</i>		<i>BX2*,rhl²/Cy</i> x <i>rhl^{KG}/Cy</i>		<i>BX2*,cuff⁰⁰³⁷/Cy</i> x <i>cuff^{N11M25}/Cy</i>		<i>BX2*,cuff⁰⁰³⁷/Cy</i> x <i>cuff^{N11M25}/Cy</i>		<i>BX2*,zuc^{10M27}/Cy</i> x <i>zuc^{0G63}/Cy</i>		<i>BX2*,zuc^{10M27}/Cy</i> x <i>zuc^{0G63}/Cy</i>	
Female genotype	<i>BX2*,rhl²/Cy</i>		<i>BX2*,rhl²/rhl^{KG00910}</i>		<i>BX2*,cuff⁰⁰³⁷/Cy</i>		<i>BX2*,cuff⁰⁰³⁷/cuff^{N11M25}</i>		<i>BX2*,zuc^{N11M27}/Cy</i>		<i>BX2*,zuc^{N11M27}/zuc^{0G63}</i>	
Dmel_matched	16 245 668	100.00%	11 825 256	100.00%	13 389 291	100.00%	13 685 445	100.00%	20 218 897	100.00%	15 281 858	100.00%
miRNAs	3 391 308	20.88%	2 750 213	23.26%	3 478 815	25.98%	4 264 109	31.16%	3 535 506	17.49%	3 707 069	24.26%
tRNAs	563 440	3.47%	381 717	3.23%	249 630	1.86%	1 047 588	7.65%	970 824	4.80%	973 116	6.37%
ncRNAs	274 114	1.69%	250 716	2.12%	84 967	0.63%	197 184	1.44%	145 664	0.72%	319 660	2.09%
miscRNAs	1 940 438	11.94%	4 441 776	37.56%	794 777	5.94%	3 122 174	22.81%	3 326 698	16.45%	6 507 497	42.58%
Transposons	7 247 574	44.61%	2 346 980	19.85%	6 625 525	49.48%	3 077 762	22.49%	9 174 921	45.38%	2 454 899	16.06%
Introns	813 565	5.01%	377 459	3.19%	558 367	4.17%	336 092	2.46%	968 061	4.79%	287 434	1.88%
Transcripts	542 936	3.34%	875 963	7.41%	569 531	4.25%	1 145 556	8.37%	440 817	2.18%	425 596	2.78%
Intergenic	1 321 210	8.13%	320 193	2.71%	881 608	6.58%	367 902	2.69%	1 475 029	7.30%	435 582	2.85%
Remaining	151 083	0.93%	80 239	0.68%	146 071	1.09%	127 078	0.93%	181 377	0.90%	171 005	1.12%
Effective depth	13 741 790		7 001 763		12 344 884		9 515 683		15 921 375		7 801 245	
Normalisation factor	0.3172714		0.6226828		0.3531728		0.4581781		0.2738380		0.5588695	

Library ID	14		15		16		17		18		19		20		21	
Library	GRH75		GRH76		GRH27		GRH30		GRH21		GRH58		GRH88		GRH22	
Parental origin	<i>BX2*,Dcr-2¹¹¹¹/Dcr-2¹¹¹¹</i> x <i>BX2/Cy</i>		<i>BX2*/Dcr-2¹¹¹¹</i> x <i>Cy/Xa</i>		<i>P1152BX2*/Cy</i>		<i>P1152BX2*/Cy</i>		<i>P1152BX2*/Cy</i>		<i>R53BX2*/Cy</i>		<i>R53BX2*/Cy</i>		<i>R53BX2*/Cy</i>	
Female genotype	<i>BX2*/Dcr2^{1111AX}G1</i>		<i>BX2*/Cy G2</i>		<i>P1152BX2*/Cy G3</i>		<i>P1152BX2*/Cy G5</i>		<i>P1152BX2*/Cy G10</i>		<i>R53BX2*/Cy G3</i>		<i>R53BX2*/Cy G5</i>		<i>R53BX2*/Cy G10</i>	
Dmel_matched	16 947 093	100.00%	15 269 297	100.00%	17 430 597	100.00%	18 065 034	100.00%	15 429 273	100.00%	16 279 059	100.00%	16 920 341	100.00%	13 871 755	100.00%
miRNAs	3 997 672	23.59%	3 473 591	22.75%	2 828 264	16.23%	2 456 132	13.60%	4 060 488	26.32%	3 916 155	24.06%	4 046 793	23.92%	2 720 348	19.61%
tRNAs	620 994	3.66%	293 115	1.92%	111 433	0.64%	178 245	0.99%	191 691	1.24%	302 025	1.86%	505 384	2.99%	147 918	1.07%
ncRNAs	124 284	0.73%	91 470	0.60%	99 510	0.57%	126 380	0.70%	134 384	0.87%	156 772	0.96%	218 565	1.29%	102 794	0.74%
miscRNAs	1 588 023	9.37%	1 013 759	6.64%	8 991 044	51.58%	9 726 925	53.84%	4 286 124	27.78%	1 480 806	9.10%	1 984 125	11.73%	6 807 534	49.07%
Transposons	8 073 990	47.64%	7 698 603	50.42%	3 947 448	22.65%	4 110 704	22.76%	5 048 740	32.72%	7 717 454	47.41%	7 325 666	43.30%	3 049 544	21.98%
Introns	755 550	4.46%	867 737	5.68%	440 087	2.52%	437 666	2.42%	513 910	3.33%	842 832	5.18%	916 638	5.42%	312 114	2.25%
Transcripts	435 636	2.57%	350 353	2.29%	231 976	1.33%	262 210	1.45%	289 041	1.87%	384 269	2.36%	400 014	2.36%	196 380	1.42%
Intergenic	1 225 750	7.23%	1 324 711	8.68%	678 534	3.89%	675 281	3.74%	797 943	5.17%	1 314 860	8.08%	1 357 357	8.02%	464 318	3.35%
Remaining	125 194	0.74%	155 958	1.02%	102 301	0.59%	91 491	0.51%	106 952	0.69%	163 886	1.01%	165 799	0.98%	70 805	0.51%
Effective depth	14 738 076		13 962 423		8 328 120		8 159 864		10 951 458		14 496 228		14 430 832		6 916 303	
Normalisation factor	0.2958241		0.3122579		0.5235128		0.5343076		0.3981093		0.3007594		0.3021224		0.6303769	

Library ID	22		23		24		25		26		27		28	
Library	GRH84		GRH85		GRH23		GRH29		GRH31		GRH17		GRH97	
Parental origin	P1152+RS3 BX2 ⁺ /Cy		P1152+RS3 BX2 ⁺ /Cy		P1152+RS3 BX2 ⁺ /Cy		LK-PIA BX2 ⁺ /Cy		LK-PIA BX2 ⁺ /Cy		RS3 x Canton ^y		P-1152 x w ¹¹¹⁸	
Female genotype	P1152+RS3 BX2 ⁺ /Cy ^{G3}		P1152+RS3 BX2 ⁺ /Cy ^{G5}		P1152+RS3 BX2 ⁺ /Cy ^{G10}		LK-PIA BX2 ⁺ /Cy ^{G3}		LK-PIA BX2 ⁺ /Cy ^{G5}		RS3/+		P-1152/+	
Dmel_matched	15 823 878	100.00%	50 523 384	100.00%	11 186 364	100.00%	21 593 154	100.00%	18 827 290	100.00%	20 591 093	100.00%	9 737 043	100.00%
miRNAs	2 503 425	15.82%	9 036 497	17.89%	1 603 513	14.33%	4 852 422	22.47%	3 165 469	16.81%	3 711 711	18.03%	2 074 552	21.31%
tRNAs	429 049	2.71%	1 515 203	3.00%	99 422	0.89%	128 814	0.60%	168 102	0.89%	863 587	4.19%	48 616	0.50%
ncRNAs	258 574	1.63%	757 628	1.50%	61 017	0.55%	108 460	0.50%	103 952	0.55%	122 807	0.60%	35 478	0.36%
miscRNAs	3 729 832	23.57%	9 508 301	18.82%	663 4861	59.31%	6 667 164	30.88%	9 859 094	52.37%	4 610 621	22.39%	807 308	8.29%
Transposons	6 282 282	39.70%	21 465 112	42.49%	2 072 874	18.53%	7 448 049	34.49%	4 096 768	21.76%	8 188 718	39.77%	2 184 131	22.43%
Introns	775 641	4.90%	2 496 136	4.94%	209 781	1.88%	709 770	3.29%	418 701	2.22%	1 003 326	4.87%	21 5734	2.22%
Transcripts	510 853	3.23%	1 423 633	2.82%	119 967	1.07%	336 981	1.56%	231 367	1.23%	484 108	2.35%	160 894	1.65%
Intergenics	1 201 771	7.59%	3 861 720	7.64%	338 746	3.03%	1 168 266	5.41%	676 687	3.59%	1 398 340	6.79%	338 239	3.47%
Remaining	132 451	0.84%	459 154	0.91%	46 183	0.41%	173 228	0.80%	107 150	0.57%	207 875	1.01%	81 688	0.84%
Effective depth	11 664 997		39 499 880		4 452 081		14 797 176		8 800 094		15 116 885		8 881 119	
Normalisation factor	0.3737573		0.1103769		0.9792899		0.2946425		0.4954353		0.2880898		0.4903685	

Table S1. Annotation of small RNA libraries

Small RNAs were prepared from ovaries of females of the indicated genotype (*Female genotype* row). The parental origin of these females is indicated in the *Parental origin* row (females x males). The Canton^y and the w¹¹¹⁸ strains are devoid of any transgene. Values for the different categories of sequences are expressed as their percentage of the total number of sequence reads that matched the *D. melanogaster* genome 5.49 in each library (Dm Matched row). For comparisons, libraries were normalized (normalization factor) relatively to the number of sequence reads aligning to the *D. melanogaster* genome but not to tRNAs or miscRNAs (effective depth). Library #5 has the lowest effective depth and was taken as reference to normalize the other libraries.

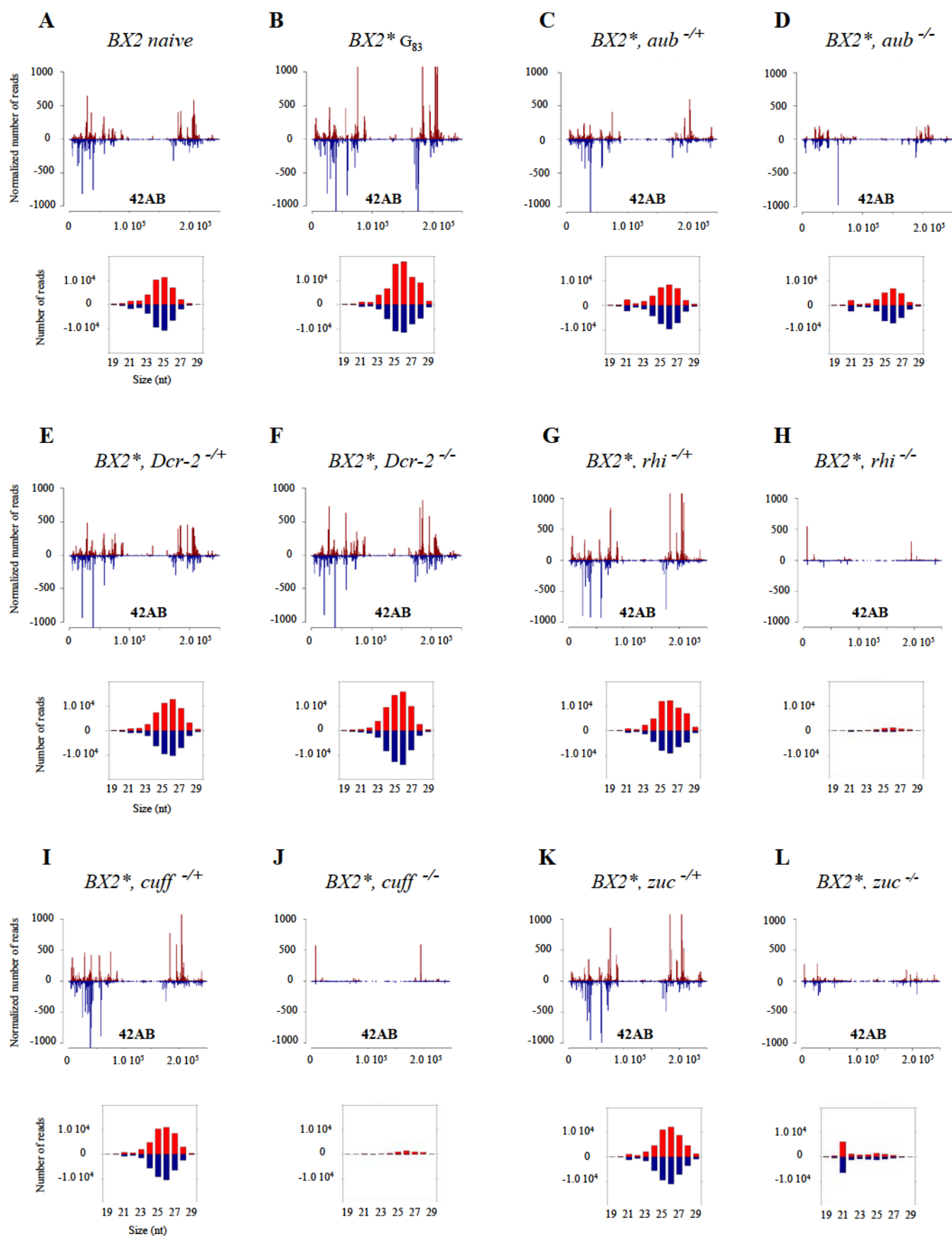


Fig. S1 – Hermant *et al.*

Figure S1. Effect of mutations affecting piRNA and siRNA pathways on small RNA production by the 42AB piRNA locus

(A-L) Ovarian small RNAs analysed in Figure 1 for *P{lacW}* homologous small RNAs were analysed for small RNAs homologous to the 42AB dual strand piRNA locus. Histograms show the length distributions of small RNAs matching the 42AB locus. Positive and negative values correspond to sense and antisense reads, respectively. Plots show the abundance of 19-29nt small RNAs. piRNA production by the 42AB locus is strongly affected by *rhi*, *cuff* and *zuc* mutations, partially affected by *aub* mutations and is not affected by *Dcr-2* mutations.

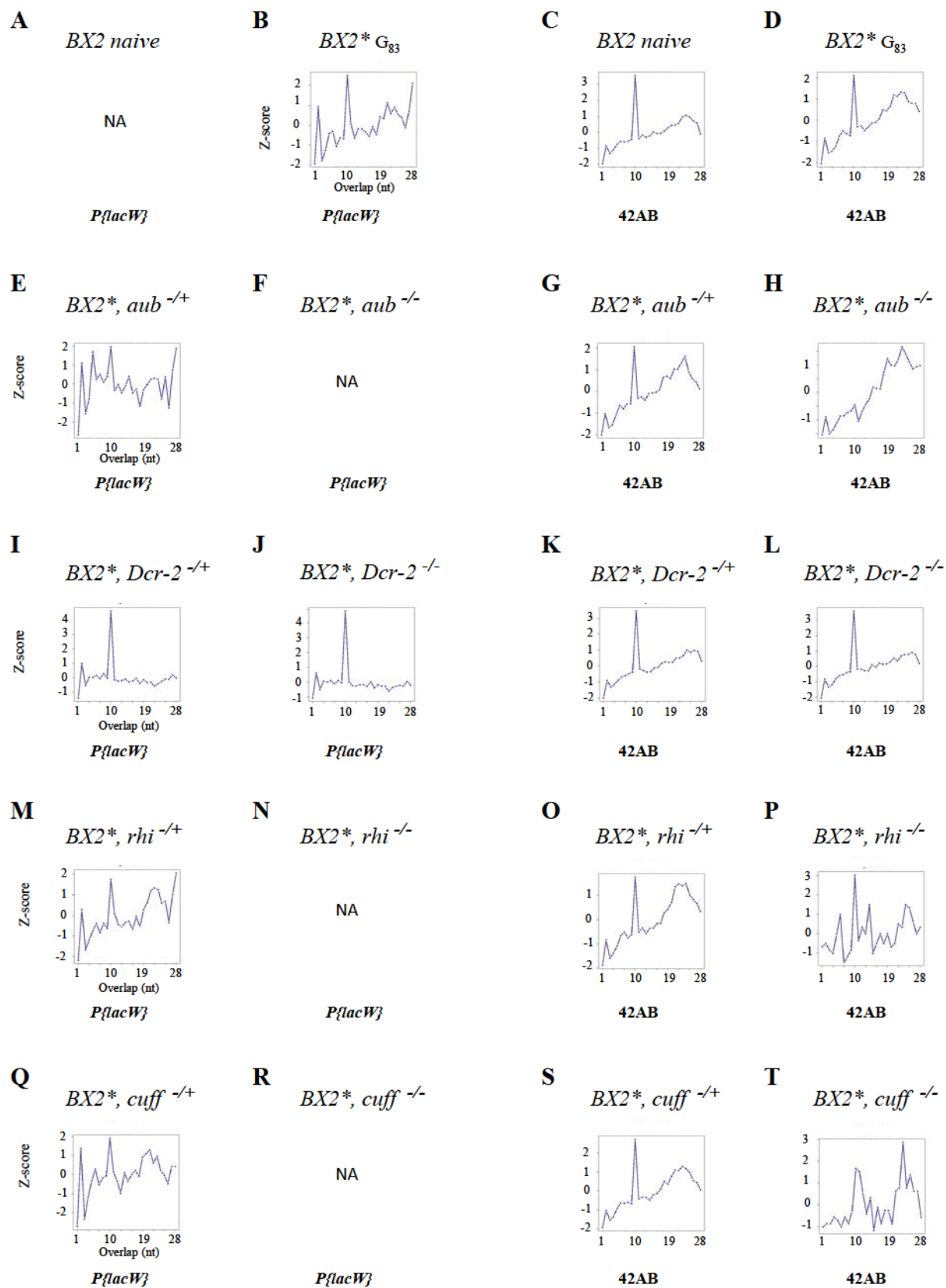


Fig. S2 Hermant *et al.*

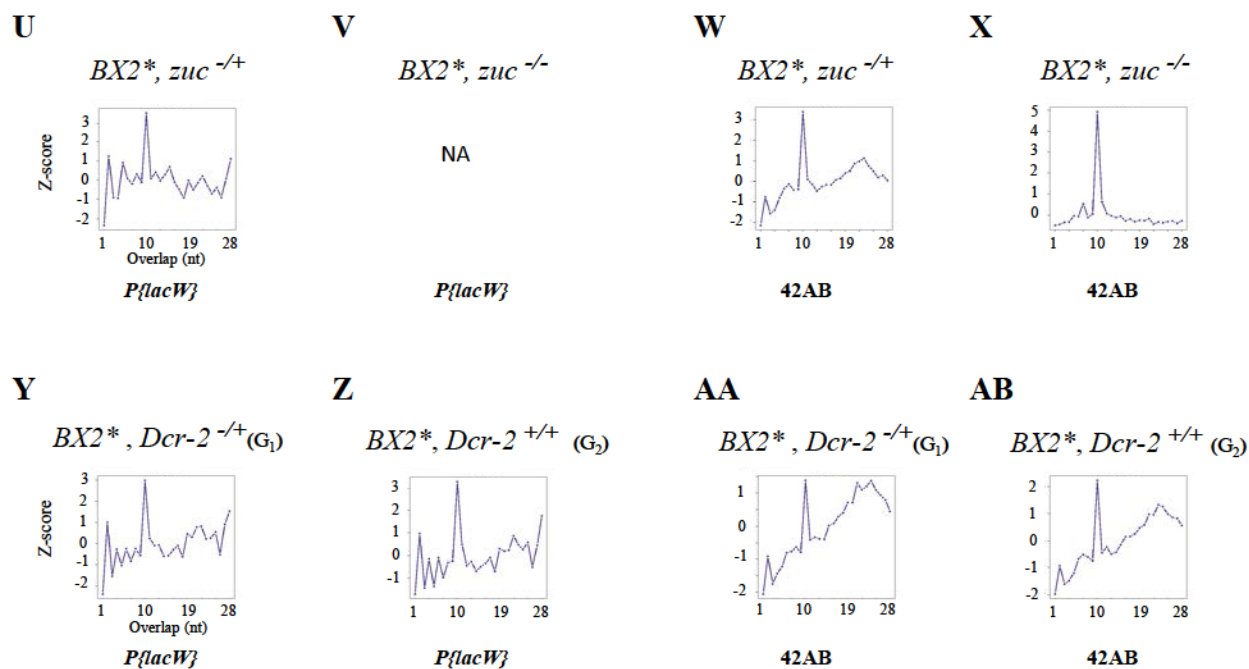


Figure S2. Effect of mutations on 23-28nt RNA ping-pong signatures

A-AB. Panels show the relative frequency (z-score) of overlapping sense-antisense small RNA pairs in the subset of 23-28nt small RNAs matching *P{lacW}* or the 42AB locus as indicated. The genotype is indicated above the graph. Y-AB correspond to G₁ and G₂ females shown in Figure 2.

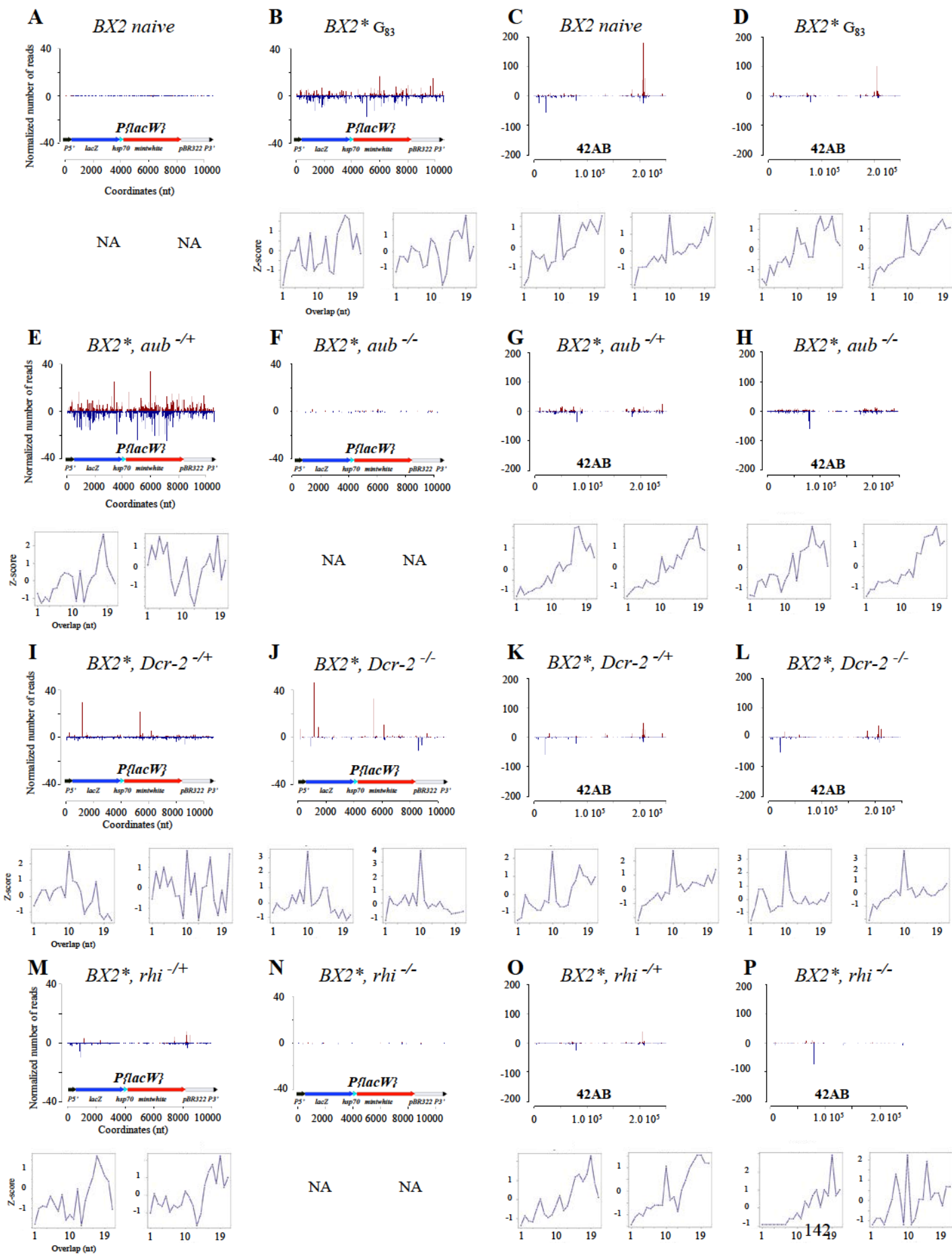


Fig. S3 - Hermant *et al.*

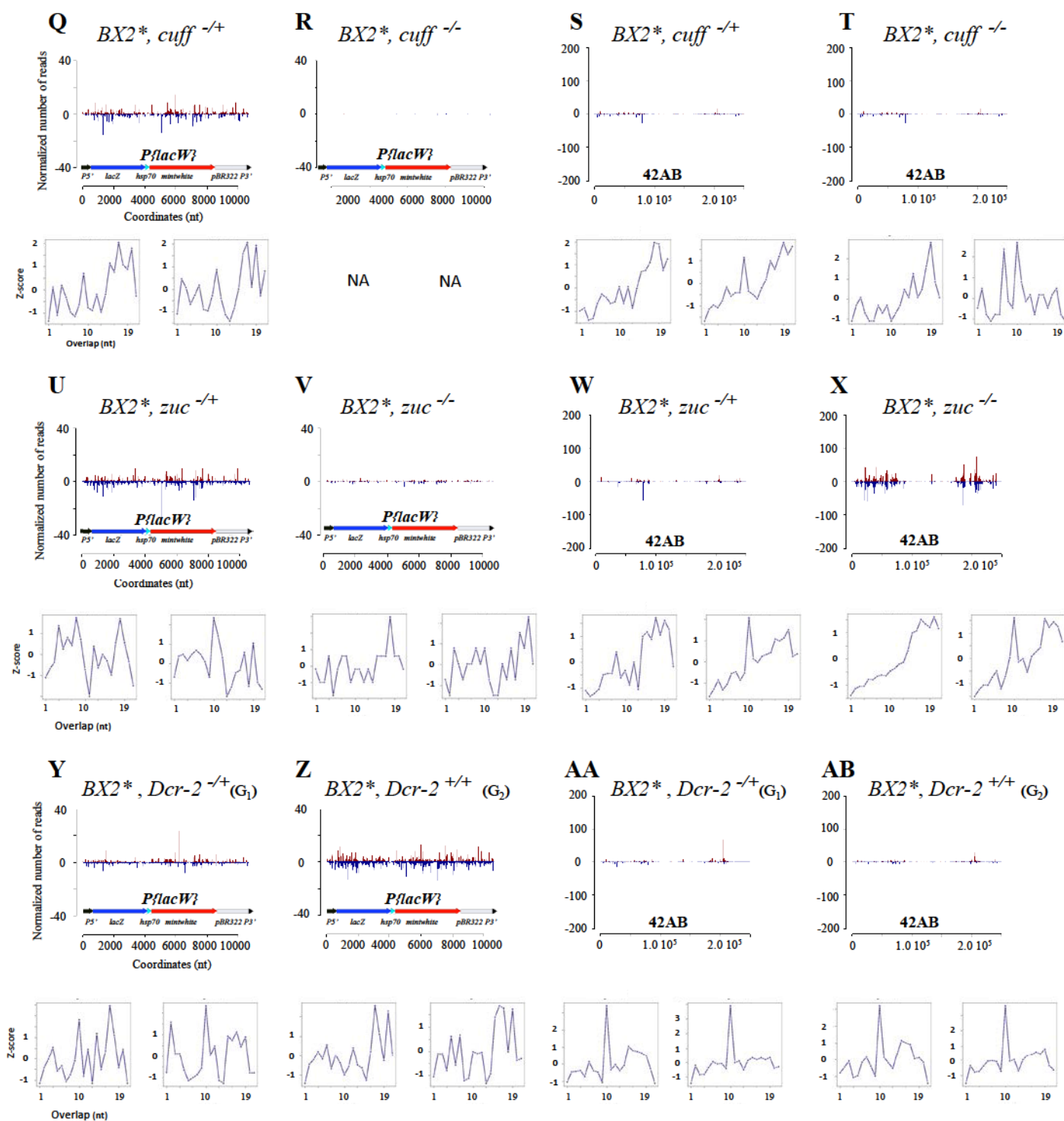
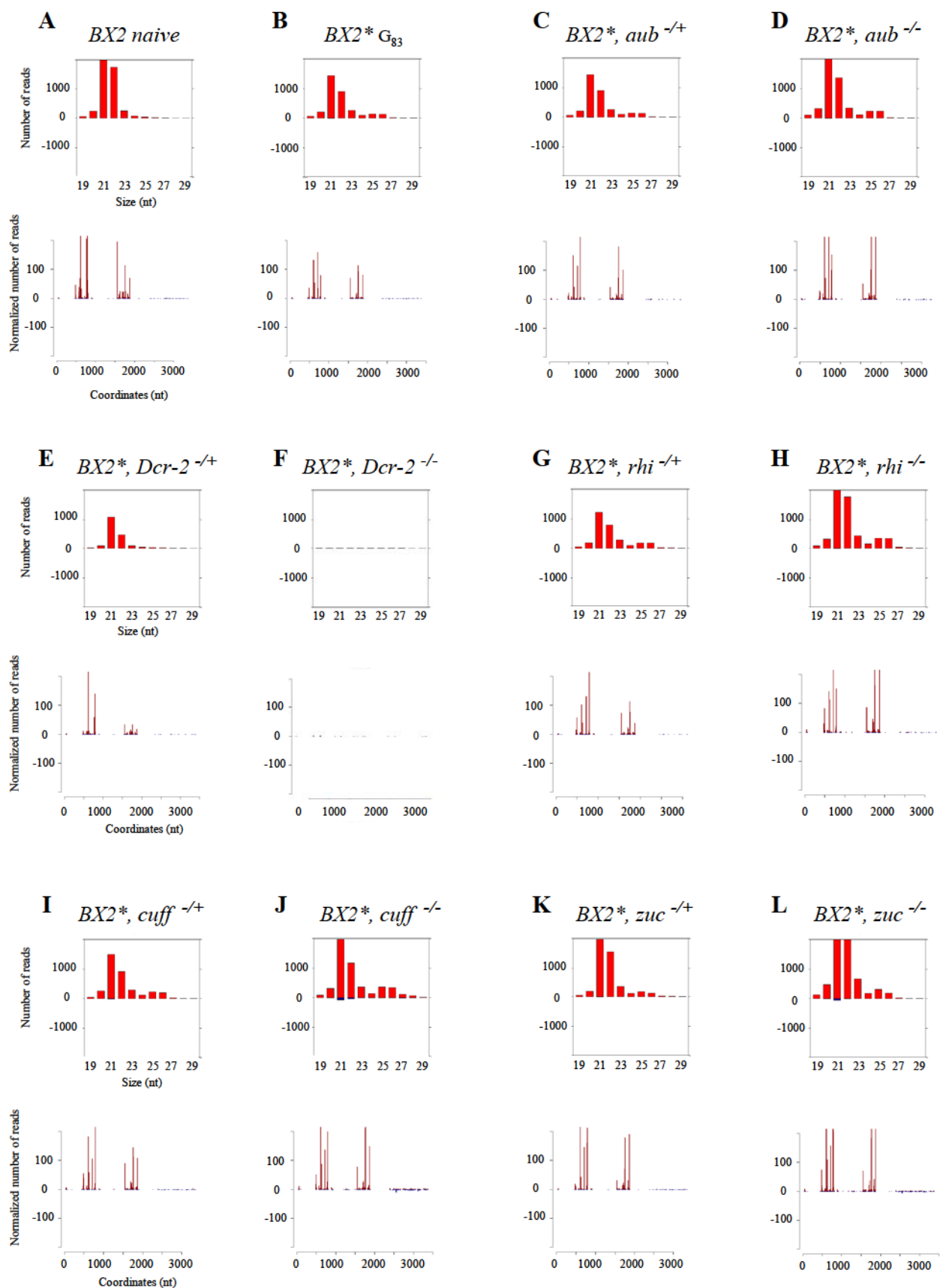


Figure S3. Effect of mutations on 20-21nt RNA abundance and ping-pong signatures

A-AB. Bar plots show the abundance of 20-21nt small RNAs matching *P{lacW}* or the 42AB locus in ovaries from females of the indicated genotype. Positive and negative values correspond to sense and antisense reads, respectively. Panels show the relative frequency (z-score) of overlapping sense-antisense small RNA pairs in the subsets of 20-21nt small RNAs (left) and of 20-21nt RNAs that pair with the indicated overlap with 23-28nt small RNAs (right), matching *P{lacW}* or the 42AB locus. Y-AB correspond to G₁ and G₂ females shown on Figure 2.



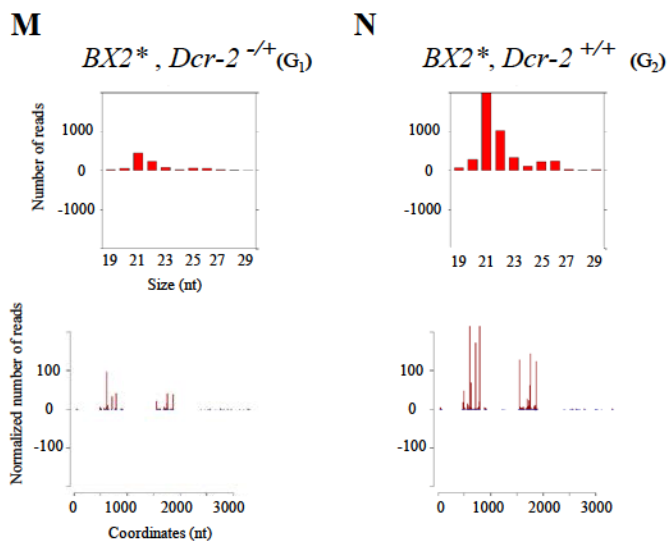


Figure S4. Effect of mutations on ovarian endo-siRNA production

A-N. Length distribution of 19-29nt small RNAs matching the *esi-1* endosiRNA-producing locus. Bar plots show the abundance of 20-21nt small RNAs on the *esi-1* locus. Positive and negative values correspond to sense and antisense reads, respectively. Production of ovarian *esi-1* 20-21nt RNAs is highly sensitive to mutants affecting *Dcr-2* but is not sensitive to mutants affecting the piRNA pathway.

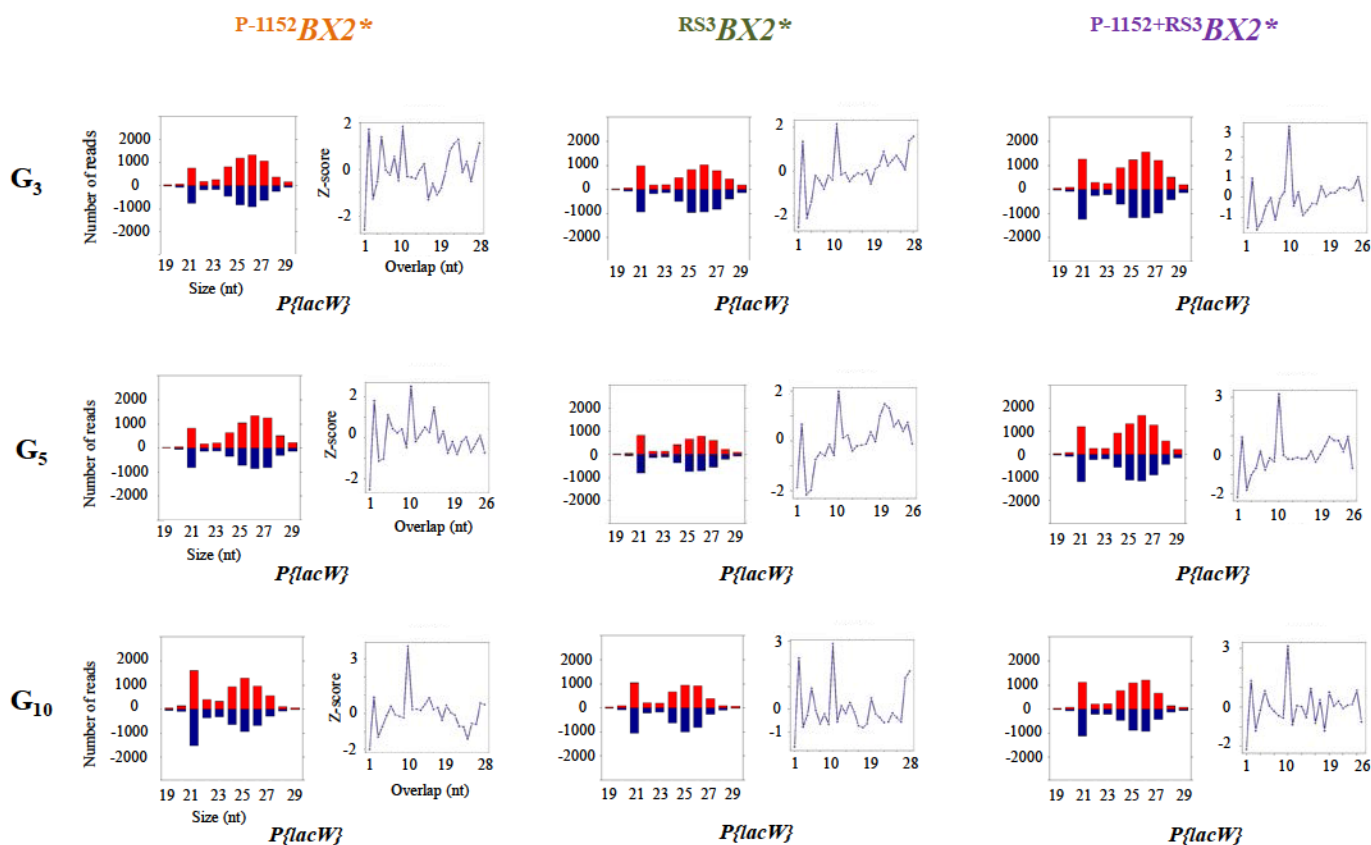


Figure S5. piRNA distribution on partially homologous paramutations

The nature of the cytoplasm inherited in G_1 females is indicated at the top of the columns. The generation at which the test was conducted is indicated on the left of the rows. **(Graphs to the left)** Length distributions of small RNAs matching $P\{lacW\}$. Positive and negative values correspond to sense and antisense reads, respectively. **(Graphs to the right)** Relative frequency (z-score) of overlapping sense-antisense small RNA pairs in the subset of 23-28nt small RNAs matching $P\{lacW\}$.

A

		TSE percentage
<i>Lk-P(1A)</i>		100% (n=1400)
<i>BX2</i> naive		0% (n> 2000)
^{T-1} <i>BX2</i> *		100% (n> 2000)
<i>Lk-P(1A)</i> <i>BX2</i>	G₃	0% (n> 2000)
	G₅	0% (n> 2000)
	G₇	0% (n> 2000)
	G₁₁	0% (n> 2000)
	G₂₆	0% (n> 2000)

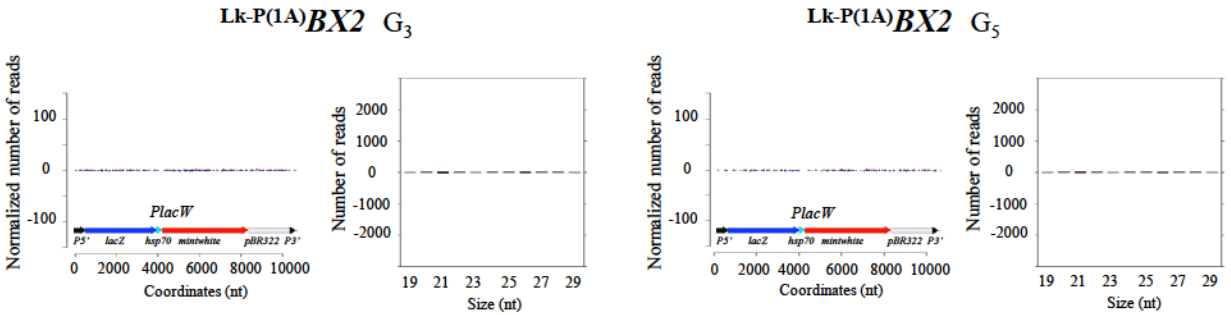
B

Figure S6. Effect of maternal transmission of piRNA, limited to the *P{lacW}* 820bp *P*-element sequence, on a *BX2* naive cluster

BX2 males carrying a *P{lacW}* naive cluster were crossed to females *Lk-P(1A)* which carry two complete *P*-elements inserted in a telomeric piRNA-producing locus at the heterozygous state. *Lk-P(1A)* females have a very strong capacity to repress *P*-element activity. Female progeny were recovered which carry the *BX2* locus but do not carry the telomeric *P*-copies. These *G*₁ females inherited a cytoplasm carrying piRNA homologous to the complete *P*-element, which therefore covers only the first 587bp and the last 233bp of the *BX2* *P{lacW}* transgene sequence (10691bp). Lines were established and the silencing capacities of these putative ^{Lk-P(1A)}*BX2** lines were tested in subsequent generations using *BQ16*. (A). In addition, deep sequencing of ovarian small RNAs was performed in *G*₃ and *G*₅ (B). TSE assays show that the silencing capacities of ^{Lk-P(1A)}*BX2* lines are null at all generations tested. Accordingly, no *P{lacW}* homologous piRNAs were detected at generations 3 and 5. Inheritance of piRNAs homologous to 820bp out of the 10691bp of the *P{lacW}* transgene does not result in *BX2* paramutation.

Conclusions

Cette étude a permis de montrer que la paramutation dépend des gènes *Rhino*, *Cutoff*, *Zucchini* et *Aubergine* qui sont impliqués dans la production des piRNA. Suite à la paramutation du cluster *P-lacW* les piRNA générés *de novo* dépendent donc de la voie de biogenèse « classique » de ces petits ARN. Par ailleurs, les siRNA produits lors de cette paramutation sont Dicer-2 dépendants mais ne sont pas nécessaire pour l'établissement du phénomène de conversion épigénétique. En effet, la paramutation est possible sans les siRNA. Par ailleurs, un apport de piRNA correspondant à certaines parties de la séquence du cluster *P-lacW* permet son activation ainsi que de la production de piRNA sur toute la longueur du transgène dès la génération G3 de façon stable au cours des générations. Enfin, cette étude met en évidence un impact de l'héritage des piRNA maternels pour la répression du cluster *P-lacW* dans les cellules somatiques.

4. Article III

Présentation de l'article III

Une précédente étude au laboratoire avait permis de montrer que les capacités de répression (TSE) d'un transgène *P-lacZ* insérés dans le locus TAS sont complètement acquises au bout de six générations quand celui-ci est apporté par voie paternelle (Josse et al, 2007). Nous proposons ici d'étudier ce phénomène d'un point de vue moléculaire en suivant la production des piRNA parallèlement à l'établissement du TSE. Ces expériences permettront d'analyser le modèle d'étude selon lequel quand un élément s'insère dans un locus producteur de piRNA des mécanismes impliquant le complexe Rhino/Deadlock/Cutoff (RDC). Celui-ci assure la transcription non canonique des dual-strand clusters par le blocage de l'épissage des transcrits (Mohn *et al.* 2014). Pour cela, nous avons joué sur les propriétés épigénétiques du TSE, l'effet maternel.

Article III

Reprogramming of a sequence fate after its insertion into a piRNA cluster

Reprogramming of a sequence fate after its insertion into a piRNA cluster

Amna Asif-Laidin¹, Valérie Delmarre¹, Antoine Boivin¹, Pauline Marie¹, Stéphane Ronsseray¹ and Laure Teyssset¹

¹Sorbonne Universités, Université Pierre et Marie Curie (UPMC), CNRS, Institut de Biologie Paris-Seine (IBPS), Unité de biologie du Développement, "Epigenetic Repression and Mobile DNA," F-75005, Paris, France.

Email: laure.teysset@upmc.fr

Running title:

Adoption of a new sequence by a piRNA cluster

Key words: transposons, piRNA cluster, Subtelomeres, Drosophila, RDC complex

Abstract

Transposable Elements (TE) are mobile elements that are repressed in germline by the PIWI interacting RNA (piRNA). In *Drosophila melanogaster*, 140 piRNA clusters are accountable for the piRNA production in ovaries. In some occasion, the genome is facing new TE invasion. In this report, using a transgenic system that mimics the *de novo* arrival of a TE in a piRNA cluster, we have studied the kinetic of production of piRNA during the first generations following this insertion. We found that the piRNA production directed against the new sequence inserted in the piRNA cluster is progressive over generations and requires three to four generations to reach an equilibrium. During this time window, splicing started to be repressed and few overlapping piRNA transcripts can be detected. These events may correspond to the signature of the recruitment of the Rhino/deadlock/Cutoff complex that defines dual-strand piRNA clusters. Therefore, the progressive efficiency of a new piRNA cluster to repress might have an impact on TE evolution when entering in a naive genome offering them opportunity to amplify before being completely repressed. We also found that production of a piRNA cluster can be modulated depending on its maternal inheritance before reaching an equilibrium state.

Introduction

Transposable Elements (TE) are mobile genetic elements that can transpose as long as they are not repressed. The most efficient mechanism known to date to regulate TE transposition in the metazoan germline is based on specialized small RNA associated to PIWI proteins, the piRNA (23-29nt) (BRENNECKE *et al.* 2007; GUNAWARDANE *et al.* 2007). In *Drosophila melanogaster*, these piRNA are produced in the ovaries from about 140 piRNA loci, called piRNA clusters, transcribed as uni- or dual-strand transcript. These piRNA clusters contain TE fragments that come from ancestral insertions of elements that have been trapped and serve as a library of euchromatic copies to repress. A huge amount of effort has been made to describe the State of Art of the piRNA biogenesis and in particular to characterize the actors of the repression for steady state piRNA clusters (THERON *et al.* 2014; CZECH AND HANNON 2016; HIRAKATA AND SIOMI 2016). Especially, it has been shown how the HP1 homolog Rhino associated with Deadlock and Cutoff, forming the RDC complex, licenses a dual-strand piRNA cluster thanks to noncanonical read through transcription, by protecting the piRNA cluster transcripts from degradation, and by blocking splicing and transcription termination (MOHN *et al.* 2014; ZHANG *et al.* 2014). However little is known about the establishment of the repression for a given TE at the beginning of the process that is to say when this new TE enters in a genome and eventually lands into a piRNA cluster. The question could be difficult to answer when considering that a given TE could be found in several piRNA clusters, as well as in diverse euchromatic loci. Moreover, because of euchromatic copies involved in the ping-pong loop of piRNA production in germline, it turns out to be complex to uncouple the dynamic of a piRNA cluster by itself.

P element is the unique element that researchers were lucky to witness its natural invasion of the *D. melanogaster* genome during the XXth century and the *D. simulans* genome at the beginning of the XXIst century (ANXOLABEHERE *et al.* 1988; KOFLER *et al.* 2015). The invasion of the *D. melanogaster* genome was successful because newly invaded *P* element appears to have a hot spot of insertion in subtelomeric regions called TAS locus (RONSSERAY *et al.* 1989), especially in the *TAS-X* of the *X* chromosome or *TAS-2R* and *TAS-3R* on the 2nd and 3rd chromosome respectively that happened to be piRNA clusters (BRENNECKE *et al.* 2007; ASIF-LAIDIN *et al.* in preparation). Therefore, when a *P* copy inserts in *TAS-X*, *P* element-piRNA are produced in female germline and euchromatic *P* elements repression is established (BRENNECKE *et al.* 2008). Alike, enhancer trap *P*-transgenes encoding *lacZ* inserted in *TAS-X* (for instance, *P[lArB]* in line *P-1152w^c*) have been shown to be able to repress female germline expression of homologous *P-lacZ* transgenes (ROCHE AND RIO 1998; RONSSERAY *et al.* 1998; DUFOURT *et al.* 2014). This repression was called *Trans-Silencing Effect* (TSE) and involves piRNA biogenesis (JOSSE *et al.* 2007; JOSSE *et al.* 2008; TODESCHINI *et al.* 2010; DE VANSSAY *et al.* 2012; MUERDTER *et al.* 2012). Using combination of

transgenes, it has been shown that exogenous sequence of the *D. melanogaster* genome (*E. coli lacZ* gene) inserted in *TAS-X* is able to repress euchromatic *lacZ* germline expression of this sequence using piRNA biogenesis pathway (POYHONEN *et al.* 2012). It has been also established that the level of TSE for a paternally inherited telomeric *P[lArB]* is low at the first generation, consistent with a maternal effect whose memory can persist for more than five generations (JOSSE *et al.* 2007). As a matter of fact, a parallel can be made between the telomeric *P[lArB]* and a newly inserted sequence in a piRNA cluster.

Therefore to address our question about the molecular mechanisms that are responsible for the erasure of the identity of a sequence whenever it inserts into a piRNA cluster, we used *P*-transgenes. They provide the unique opportunity to precisely control the genome as well as the epigenome. The genome, because we possess strains with or without the *TAS-X* piRNA cluster containing or not *P* or *P*-transgene inserted in the *TAS-X* (for instance *P[lArB]*) and strains completely devoid of *P* element. The epigenome, because piRNA against *P*-transgene can be inherited only when it is brought by the mother.

Therefore, by using *P*-transgenes we can control maternally inherited piRNA that are able to establish and maintain heterochromatin environment across generations for piRNA clusters in order to ensure efficient repression of euchromatic transposons throughout the genome. We tested in parallel to the piRNA production, the functionality of those newly produced piRNA by measuring the TSE at each generation. We are demonstrating that while the piRNA cluster is stable across generations, the exogenous sequence, which has landed into, becomes in turn also a piRNA cluster by acquiring specific properties of this cluster and by accessing to a fully repression state progressively. It includes the production of piRNA most likely due to the presence of RDC complex, which can allow the transcription but blocks the splicing events. The mRNA produced initially by the sequence turns out into a piRNA precursor since we have observed an increase of piRNA arising from intronic and overlapping regions. Our observations are also in favor of the fact that the memory of a state at one generation could be transmitted to the next since three to four generations are necessary to establish a new functional piRNA cluster that reaches the equilibrium.

Materials and Methods

Fly stocks

P-1152^w carries two *P-lacZ* fusion enhancer trap transgene (FBti0005700) inserted in *TAS-X* that contains an in-frame translational fusion of the *E. coli lacZ* gene to the second exon of the *P* transposase gene, as well as the *rosy+* gene as a transformation marker and the *Adh/Adhr* gene (O'KANE AND GEHRING 1987) (Fig. 1). This strain contains TAS subtelomeric repeats on *TAS-X* and *TAS-3R*, and is mutant for the *white* gene. *BQ16* carries also a *P-lacZ* fusion enhancer trap transgene (FBtp0000154) located at 64C in euchromatin strongly expressing *lacZ* in germline cells. When *P-1152^w* females are crossed by *BQ16* males, *lacZ* is repressed in female germline (JOSSE *et al.* 2007). *Canton^y* derived from the *Canton-S* by a mutation in the *yellow* gene.

Experimental Conditions

All crosses were performed at 25°C and involved five couples. *lacZ* expression assays were carried out using X-gal overnight staining as described in Josse *et al.*, 2007. *Trans*-silencing was quantified as previously described by determining the percentage of egg chambers with no *lacZ* expression in the germline (TSE%) (JOSSE *et al.* 2007).

Ovarian Small RNA sequencing and analysis

Total ARN extraction, small RNA libraries preparation and sequencing were performed as described in Hermant *et al* (HERMANT *et al.* 2015). Sequence reads in fastq format were trimmed from the adapter sequence 5'-TGGAATTCTCGGGTGCCAAG-3' and matched to reference sequences using Bowtie (LANGMEAD *et al.* 2009). The 23-29nt reads (piRNA) and 21nt reads (siRNA) matching the reference sequences with 0 mismatch were retained for subsequent analysis. Annotation of small RNA libraries is described in **Supplementary Table 1**. Sequence length distributions, small RNA mapping were generated from bowtie alignments using Python and R (www.r-project.org/) scripts, which were wrapped and run in a Galaxy instance publicly available at <http://plastissipi.fr>. Tools and workflows used in this study may be downloaded from this Galaxy instance

Results

Activation of a quiescent piRNA cluster

Upon arrival of a new TE, the genome is facing an invasion that has to be repressed in order to survive. Today we are witnessing successful events leading to control their transposition that, for the most, required integration of at least one copy into at least one piRNA cluster. The newly inserted copy will eventually synthesize piRNA against its own family. The purpose of this study was to understand what happens right after a sequence integrate into an active piRNA cluster. In other words, does the sequence immediately acquire piRNA cluster properties or does it take a certain number of generations for the inserted sequence to partially or totally lose its identity and progressively acquire piRNA cluster properties?

To address this question, we first make use of *P*-transgenes inserted in *TAS-X* that give us the opportunity to precisely control the genome as well as the epigenome. The *P-1152w^c* line carries two copies of a *P[lArB]* transgene inserted into *TAS-X* (**Fig. 1**) and a mutation in the *white* gene controlling eyes pigmentation (*w*). The *Canton^y* strain derived from the *Canton-S*, a classical laboratory strain, possesses a mutation in the *yellow* gene (*y*) affecting carcasses pigmentation and is devoid of *P* element as well as the *TAS-X* locus (ASIF-LAIDIN *et al.* in preparation). The *y* and *w* marker genes were used to control the telomeric region in successive generations as well as to identify rare recombination events that could distort further analysis. Using these two strains, we try to simulate the situation of a newly inserted TE into a piRNA cluster.

In G0, on one hand five *P-1152w^c* females were crossed by five *Canton^y* males to constitute the maternal inheritance of the *P[lArB]* transgene, designated as *P[lArB]-MI* (**Fig. 2 A**). On the other hand, five *Canton^y* females were crossed with five *P-1152w^c* males to constitute the paternal inheritance of the transgene, designated as *P[lArB]-PI* (**Fig. 2 B**). Four sets of five females of both G1 hemizygous progenies were then crossed with five *Canton^y* males to constitute the G2 progenies. The following generations were obtained by the same mating scheme using hemizygotes females crossed with *Canton^y* males. The four sets of crosses with *P[lArB]-PI* were named lines A, B, C, and D and those with *P[lArB]-MI* were named lines E, F, G, H. To measure the repression capacity of the telomeric transgene in each experiment, we used a functional test called *Trans-Silencing Effect* (TSE) that is based on the capacity of a telomeric *P-lacZ* transgene to produce piRNA that repress germline expression of a homologous euchromatic *P-lacZ* reporter transgene (JOSSE *et al.* 2007). TSE corresponds to the percentage of *lacZ* repressed egg chambers. It was measured at each generation up to the ninth by crossing five females of each subline at each generation with five males containing a *P-lacZ* reporter transgene (*BQ16*). Results show that the *P[lArB]-MI* is repressing strongly the

P-lacZ reporter transgene with an average of about 88% of TSE over generations for the four sublines (**Fig. 2 C and Supplementary Table 2**). On the contrary, *P[lArB]-PI* is establishing only a weak TSE at G1 (9.25%), and then increasing rapidly in the second generation (41% for the weakest subline) to reach the plateau at the fourth generation (>90%) (**Fig. 2 C and Supplementary Table 2**). Thus a progressive reinforcement is observed as the number of generations increases. Such partial epigenetic transmission through meiosis has already been observed with a different set of transgenes (JOSSE *et al.* 2007). For practical purposes, at that time the kinetic was performed in presence of the reporter transgene. In the present study, the kinetic seems faster considering that the maximum of repression is reached in less generations than the six generations that were previously observed. Such differences could be simply due to different combination of transgenes and genetic backgrounds. Nevertheless, the *P[lArB]-PI* does not acquire full functionality from the first generation but instead need few generations (less than four) to acquire progressively full repression capacity. Moreover, these results bring out that the contribution of euchromatic target in the establishment of a stable repression could be ruled out.

Based on these results, it could be hypothesized that the progressive capacity to repress the euchromatic *P-lacZ* target by the *P[lArB]* inserted in *TAS-X*, could be the picture of a progressive acquisition of piRNA biogenesis from the *TAS-X* piRNA cluster itself across the generation in the *P[lArB]-PI*. The *TAS-X* locus is a particular piRNA cluster in the sense that it could be present or absent without harming laboratory strains (ASIF-LAIDIN *et al.* in preparation). This locus is made of several 1.8kb long non coding repeats, called *TAS-X* repeats (KARPEN AND SPRADLING 1992). Consequently, strains with (*i. e.* *P-1152w^c*) or without (*i.e.* *Canton^y*) *TAS-X* locus at the tip of the *X* chromosome are available (ASIF-LAIDIN *et al.* in preparation). The *TAS-X* repeats shares homology with *TAS-2R* and *TAS-3R* located on the 2nd and 3rd chromosomes, as well as *INV-4* solo LTR sequences dispersed in the genome. To investigate the dynamic of the ovarian piRNA production by the *TAS-X* locus, we conducted an analysis using a domain that is specific and unique for the *TAS-X* locus, the 909bp domain F (ASIF-LAIDIN *et al.* in preparation). For now on, we will assume that this domain is a representative of the behavior of the whole *TAS-X* repeats. To look at the distribution of piRNA along this domain F for the *P[lArB]-MI* and *P[lArB]-PI*, we sequenced ovarian small RNA from successive generations: from the first (G1) to the fourth (G4) generation for sublines H having a *P[lArB]-MI* and D having a *P[lArB]-PI* and from G1 to G3 for subline B having a *P[lArB]-PI* in heritage. The 23-29 nt domain F-piRNA identified in each libraries were plotted on domain F sequence. As expected, for the *P[lArB]-MI*, the quantity and distribution of these piRNA reads appear homogeneous across the domain F sequence with conserved discrete pics for the *P[lArB]-MI* (**Fig. 3A, Supplementary Table 3A and Supplementary Fig. 1**). For the study of the paternal lineage, we first verified that the *Canton^y* strain does not produce domain F piRNA (**Supplementary Table 4**). In a sense, *Canton^y* strains

can be considered as a G0. Surprisingly, the domain F *P[lArB]-PI* for both sublines B and D acquires in only one generation a complete capacity to produce piRNA (**Fig. 2D, 2H and Fig. 3**). Once acquired this quantity is stable across generations (**Fig. 3**). As a control strain, we also analysed domain F piRNA synthesized in ovaries of the *P-1152w^c* strains that has a homozygous *TAS-X* locus. The amount of piRNA produced by the piRNA cluster decreased by about half between a homozygous locus (7477 normalized reads) and heterozygous locus (4252 normalized reads).

Therefore, contrary to what was expected, while the repression capacity of *P[lArB]* transgene inserted in *TAS-X* is constant across generation for the *P[lArB]-MI* inheritance and increasing across generation for the *P[lArB]-PI*, the piRNA cluster where this transgene is inserted is producing the same amount of piRNA from G1 to at least G4. This discrepancy might suggest that the kinetics of *P[lArB]* piRNA production is different from the *TAS-X* piRNA production during the first generations. For the *TAS-X* cluster, we propose that piRNA synthesized by the *TAS-3R* and/or *TAS-2R* and at least the 42AB cluster that possesses a fragment of *INV-4* present in the autosomes of *P-1152w^c* and *Canton^v* strains are involved in a transgenerational epigenetic conversion that results in an activation of the *TAS-X* locus, as it was previously described for a tandem repeat of seven *P-lac-white* cluster (DE VANSSAY *et al.* 2012). Such stimulation of piRNA production could not occur on the *P[lArB]* transgene in the *P[lArB]-PI*, because no homologous sequences capable of producing maternal deposited piRNA are present in G0.

Differential piRNA kinetic between the *P[lArB]* and the *TAS-X* cluster

Therefore the *P-1152* transgene system is mimicking what is happening right after a new TE lands into a piRNA cluster. The new TE insertion, as the *P[lArB]* in the *P[lArB]-PI*, is surrounded by maternally deposited piRNA targeting the flanking piRNA cluster. The question is what is happening to the TE expression? To directly address this question, we analyzed the 23-29 nt piRNA synthesized by the *P[lArB]* transgenes in both inheritages.

For the subline H having a *P[lArB]-MI*, the amount of piRNA against the full length of *P[lArB]* appears about constant across generations (from G1 to G4) and are homogeneously distributed all along the sequence even so a dense region of piRNA in the region annotated "Hsp70" is observed (**Supplementary Table 3A and Supplementary Fig. 1**). This region corresponds to two convergent transcriptional terminators from the *Hsp70Ab* and *CG3281* genes. The higher amount of piRNA from this region of the transgene could be correlated to the fact that some piRNA corresponding to the region are produced in the *Canton^v* strain (**Supplementary Table 4**) and that the *CG3281* gene appears to be an endogenous piRNA sources (**Supplementary Fig. 7**) as it was already described in a previous study for the entire Hsp70 locus (MOHN *et al.* 2014). This result presenting evidence that at the equilibrium state the *P[lArB]* transgene inserted in *TAS-X* is producing piRNA is not new (TODESCHINI *et al.* 2010; DE VANSSAY *et al.* 2012; MUERDTER *et al.* 2012; LE THOMAS *et al.* 2014). Here we show that the transcript is sliced with a reproducible pattern across generations by the piRNA machinery in ovaries. For the *P[lArB]-PI* transgene, the number of piRNA corresponding to the telomeric transgenes in the two sublines gradually increased with the number of generations (**Fig. 2F, 2H, Supplementary Table 3 and Supplementary Fig. 3**). To directly compare the evolution of piRNA quantity synthesized at each generation in both lineages, we analyzed the normalized *P[lArB]*-piRNA reads in *P[lArB]-PI* and *P[lArB]-MI*. As mentioned above, we included piRNA reads found in the *P-1152^{w^c}* library as representative of G0 for a maternal inheritance because it carries homozygous *P[lArB]* transgenes, as well as piRNA extracted from *Canton^v* as representative to the G0 for a paternal inheritance because it is devoid of *P[lArB]* transgenes (**Supplementary Table 4**). Analysis of **Fig. 3B** shows that, as it has been already observed, the amount of piRNA is reduced between the homozygous and the heterozygous *P[lArB]* transgenes-containing strains (DE VANSSAY *et al.* 2012; LE THOMAS *et al.* 2014). In this study, we show that whereas the repression capacity of *P[lArB]* is stable across generations (see % of TSE in **Supplementary Table 2**), the *P[lArB]* is progressively reducing its capacity to produced piRNA when it is transmitted heterozygously in the *P[lArB]-MI* to end up with half of piRNA after three generations (**Supplementary Table 3, Supplementary Table 4 and Fig. 3**). By contrast, the *P[lArB]-PI* is progressively acquiring the capacity to produce piRNA and this is correlated to an increasing

capacity to repress the *P-lacZ* reporter. Interestingly, the emergence of *P[lArB]-PI* is carrying out with the same rate than the *P[lArB]-MI* descent, in about four to five generations, to end up at the same level of magnitude of piRNA production coming for both lineage. The equilibrium status is reached for both lineages for the *P[lArB]* transgene and *TAS-X* locus but with a totally different kinetics. Therefore, these results are showing a direct correlation between number of piRNA and repression efficiency for the *P[lArB]* transgene paternally inherited, as well as the dynamic of piRNA production that can support variation in the number of piRNA and still being capable of strong repression.

The *P[lArB]* transgene contains distinct sequences from the *Drosophila* genome: *Adh/Adhr* and *rosy* genes, convergent transcriptional terminator of the *Hsp70/CG3281* region and exogenous sequences (*lacZ* and *pBluescript*). We then wondered if all these sequences respond as the whole transgene does or if they show different behavior with regard to emergence of their piRNA production capacity. **Supplementary Fig. 1, 2 and 3** show that piRNA reads are distributed throughout each sequences for the *P[lArB]-MI*, while number of piRNA reads matching to the sequences gradually increases across the generations for the *P[lArB]-PI*. Together with **Supplementary Table 3** and consistent with the whole *P[lArB]* transgene, the number of piRNA reads for each region decreased in the *P[lArB]-MI* and increased in the *P[lArB]-PI*. Graph in **Fig. 2H** emphasizes that all type of sequences, even the exogenous (*lacZ* and *pBluescript*), behave the same and accumulate the capacity to produce more and more piRNA over generations to reach a level close to the maternal inheritance pattern after four generations (ratio close to 1). This result proves that exogenous element, as new TE for instance, can be handled like endogenous sequence and if this sequence lands into a piRNA cluster, piRNA will be produced efficiently all across the sequence in three to four generations.

Therefore, propagation of piRNA production all along the body of the transgene occurs. But it requires few generations for the transgene to acquire similar properties to the flanking piRNA cluster. We hypothesize that the time needed to acquire the complete capacity to repress corresponds to the time needed by the RDC machinery to fully take place and initiate the switch from an inactive to an active *P[lArB]* piRNA production. To investigate the kinetic of piRNA production of the *TAS-X* locus itself, we analyzed the small RNA libraries and found that the piRNA do not follow the same kinetic as the telomeric transgene. Instead, the synthesis is maximum at the first generation and then stays stable. This discrepancy of kinetics between *TAS-X* locus and *P[lArB]* demonstrates a different status. For the *TAS-X* cluster, piRNA synthesized by the *TAS-3R* and/or *TAS-2R* present in the *P-1152w^c* and *Canton^v* strains are activating piRNA production by the *TAS-X* locus in one generation (DE VANSAY *et al.* 2012; ASIF-LAIDIN *et al.* in preparation). Therefore, this experience mimicking the kinetic of production of piRNA by a sequence newly inserted in a piRNA cluster

revealed a specific dynamic of the sequence at the beginning of the process that differs from the activity of the piRNA cluster.

RDC on the *P[lArB]* transgene: From mRNA to piRNA production

The current model proposes that nuclear Piwi loaded piRNA target dual-strand clusters that allow deposition of H3K9me3 marks on histones. These histones are then targeted by the HP1 homolog Rhino that will recruit, via Deadlock, the Cutoff factor. Cutoff by binding the 5' uncapped piRNA cluster transcripts will stabilize them and block splicing and termination of transcription. These steps initiate piRNA production in the nuage after their UAP56 directing nuclear export (KLATTENHOFF *et al.* 2009; LE THOMAS *et al.* 2014; MOHN *et al.* 2014; ZHANG *et al.* 2014). To directly address the role of the Rhino/Deadlock/Cutoff (RDC) complex recruitment to the telomeric transgene and test if a shift of transcription to a piRNA transcription type establishes as quick as the piRNA production, we first looked at the piRNA reads mapping to introns and exons of the *Adh/Adhr* and *rosy* genes that are included in the *P[lArB]* transgene. FlyBase databases indicates that *Adh/Adhr* genes are expressed in ovaries, while *rosy* gene is not. The two loci provide us good tools to test the switch between an mRNA transcription to a piRNA cluster non-canonical transcription by looking at the piRNA reads matching to introns and exons during the process. We calculated the ratio between normalized reads synthesized by mature mRNA (exons) versus reads coming from immature mRNA (introns) normalized per kb (**Supplementary Table 5**). As previously, we also included as representative of G0 for the maternal inheritance the value calculated from the *P-1152w^c* line that carries the homozygous *P[lArB]* transgene and as G0 for the paternal inheritance we used the *Canton^s* strain devoid of transgene. When considering the *rosy* gene that is not expressed in ovaries, the ratios obtained in the *P[lArB]-MI* and *P[lArB]-PI* were approximately the same with a slight decrease going from four to more than two fold more exons than introns. For the *P[lArB]-MI Adh* expressed gene, six to eight fold more exonic piRNA reads are produced, whereas for the *P[lArB]-PI*, at least thirteen times more exonic piRNA reads are produced in G1, that drop in G2 to reach the same level of magnitude than the *P[lArB]-MI* (**Fig. 4A and Supplementary Table 5**). These results are in good agreement with what is expected for a gene inserted in piRNA cluster that will lose, when expressed, its mRNA identity to switch to a piRNA cluster identity without splicing. The *P[lArB]* transgene changes his fate in just one generation. Reinforcing that observation, overlapping piRNA between exons and introns for *Adh* and *rosy* genes were detected (**Figure 5A and B**). Nevertheless the RDC does not abolish completely the splicing since the equilibrium at G4 is not 1. The *rosy* gene was of interest to look at because it is not expressed in ovaries, therefore no potential conflicts between mRNA and piRNA cluster transcription could occurs. It appears that the amount of piRNA

per kb is lower than what was obtained for the *Adh* gene, most likely because the reads are produced only from the *TAS-X* locus whereas for the *Adh* gene, reads are probably coming from the *TAS-X* locus and from the endogenous *Adh* gene locus. Interestingly even in this configuration the RDC is not strong enough to abolish splicing events since two times more exon than introns reads are produced, suggesting that transcript coming from the piRNA cluster might be actually spliced. In agreement with previous results suggesting that the efficiency of piRNA cluster transcription increased, we are also observing an increase of piRNA that might correspond to an increase of transcription occurring between the first and the second generation (MOHN *et al.* 2014).

Therefore splicing events along the whole transgene appear to be progressively partially inhibited. We then wanted to know how transcription occurs in the *P[lArB]* transgene. In other words does an RNA molecule initiated in the *TAS-X* locus transcribes the whole transgene or are smaller molecules produced? We looked at the ratio between sense and antisense piRNA reads for the different sequences present in the transgene in the *P[lArB]-MI* and *P[lArB]-PI*. We quantified the ratio between sense and antisense piRNA by arbitrary fixing the sense strand as the one transcribed from the strand oriented from telomere towards the centromere (Supplementary Fig. 4, 5 and 6). From the Fig. 4B, it seems that there is a tendency to have a ratio close to 1 in the maternal inheritance, meaning that at the equilibrium there are as many sense as antisense piRNA synthesized from the whole transgene. Nevertheless, for the paternal inheritance, it seems that the equilibrium state is not reach yet. The imbalance between sense and antisense piRNA is more pronounced in the region containing the double transcriptional terminators. This might reflect that the inhibition of transcriptional termination is incomplete during the first generation after the insertion of a sequence in a piRNA cluster. Even if this occurs, the *P[lArB]* appears functional anyway for the production of piRNA that repress their targets.

To get further insight into the piRNA production from the *P[lArB]* transgene, we looked at the size distribution of ovarian small RNA corresponding to the sequences comprising the telomeric transgene. Additionally to the 23-29 nt piRNA, *lacZ*, *Hsp70*, *Adh*, *rosy*, and *pBluescript* show also a huge amount of 21nt siRNA compared to piRNA quantity on the first generation of the 2 *P[lArB]-PI* sublines (Fig. 2E, 2F and Supplementary Fig. 6, 7). Such phenomenon was not observed neither from the domain F of *P[lArB]-PI* nor for all of the sequence of the *P[lArB]-MI* inheritance (Supplementary Fig. 5). In the following generations, the ratio between the siRNA and piRNA decreased (Supplementary Table 7 and Fig. 4). A possible explanation will take into account that at the beginning of the process for the *P[lArB]-PI* the locus will be transcribed in both sense due to the different promoters present on the transgenes as well as on the TAS locus. This might produce double stranded RNA targeted by the siRNA machinery. Once the RDC

complex started to be recruited to the locus, piRNA started to be produced, decreased availability for the siRNA machinery. It has been suggested that the RDC increase the level of transcription of piRNA locus (MOHN *et al.* 2014). This might explain why it appears that the quantity of siRNA is stable while the quantity of piRNA is increasing in the P[lArB]-PI. At the steady state, a low level of siRNA is produced (see line P[lArB]-MI and P-1152w^c). This result also indicated that siRNA by themselves have no TSE repression capacity since in G1 the percentage of TSE is low and abundant siRNA are synthesized.

Finally, we wanted to investigate the propagation of piRNA production to the flanking region of the P[lArB] transgene and within the transgene itself. First we were able to detect only two 27nt piRNA overlapping the 3' of the P[lArB] and TAS-X locus, one in the P[lArB]-MI and one in the P[lArB]-PI G1 lineage. We then looked at overlapping piRNA between the "rosy" region and the pBluescript region. For this region we found at least twenty one and seven overlapping normalized piRNA reads in P[lArB]-MI and in P[lArB]-PI respectively. This might be in favor that a transcript is synthesized through these two regions.

Given our observations about piRNA distributed to the respective region of the P[lArB] transgene, we then looked at the impact over generations of piRNA on the genomic endogenous loci in order to detect potential propagation of piRNA production on flanking genomic sequences. The "Adh" region included in the transgene does not contain the entire *Adh* gene, missing the two last exons and contains more than 1kb in 5' of the first exon containing promoter region. The footprint of the region included in the transgene is clearly visualized on **Figure 5A** where sense and antisense piRNA are mapped even in the promoter region. Nevertheless the 3' region of *Adh* that is not included in the transgene is producing progressively more and more sense piRNA across generations for the P[lArB]-PI and in the P[lArB]-MI. These sense piRNA of 25 and 27 nt have a 1U bias (70%) and could be the results of the endogenous *Adh* mRNA degradation. Interestingly, a discrete pic of these sense piRNA is observed just after the boundary between the sequences included in the transgene and the endogenous *Adh* gene. This pic of sense piRNA could correspond to the responder piRNA, while the sense piRNA observed on the tail of the mRNA could correspond to the trailer piRNA that have been recently described (HAN *et al.* 2015; MOHN *et al.* 2015; WANG *et al.* 2015). Except these phased piRNA that are related to the transcript, we did not detect piRNA corresponding to the flanking region. The "rosy" region included in the transgene contains the full *rosy* gene, as well as the 5' truncated *l(3)87Df* gene that has the same orientation of transcription than *rosy*. Once again the footprint of sequences included in the transgene is clearly visible and increases over generations. In this locus, no propagation of piRNA production is observed on the *l(3)87Df* gene, because no phasing can occur on this strand (**Fig. 5B**). The "Hsp70" region included in the transgene contains the last exon of *hsp70Ab* gene as well as the 3' region of the *CG3281* gene (**Supplementary Fig. 7**). As observed above, the propagation of piRNA on the flanking

genomic regions occurs in the transcript and appears mostly sense to the transcript. This is in agreement with previous results that identified piRNA production from TE flanking genomic sequences (SHPIZ *et al.* 2014). Therefore we did find evidence of propagation of piRNA in flanking region in our experiences only if the flanking region is transcribed. In this case, the piRNA found were sense piRNA suggesting that there are not produced by ping-pong amplification loop.

Discussion

The comprehension of Transposable Elements regulation have made a leap forward a decade ago with the identification of the PIWI-interacting RNA, that appear to be a broad approach to control transposition in Metazoan germline. Since then, a huge amount of literature accumulates dissecting the precise molecular mechanism leading to an efficient repression (THERON *et al.* 2014; IWASAKI *et al.* 2015; CZECH AND HANNON 2016). Thanks to all this work, we are starting to apprehend the smart way of their regulation even if several shadows persist. Early on, it has been established that piRNA are synthesized from about 140 centromeric and telomeric heterochromatic piRNA clusters in the *Drosophila* ovaries. A quite remarkable point is that these 140 clusters are apparently conserved since they are for the most part conserved in laboratory strains as well as in strains recently collected from the wild. TE are known to be able to jump into other TE even if they are not belonging to the same family. These foci of TE, called nested TE, will eventually constitute heterochromatin resulting in a stack of pieces of interrupted TE (PIMPINELLI *et al.* 1995; FANTI *et al.* 1998; FRANSZ *et al.* 2000) being at the origin of some heterochromatic regions. Other TE can then continue to insert and eventually degenerate by recombination and/or deletion. In metazoan, these loci are at the origin of the piRNA clusters that correspond to libraries of trapped sequences which autoregulate their euchromatic copies. Therefore, these trapped TE can be considered as domesticated sequences that fulfill the function of guardian of their host genome integrity.

But what is happening when a new element entered in a naive genome? That is to say when no trapped copies able to establish the epigenetic repression based on piRNA are inserted in one of the piRNA clusters. If the element is transposing all over the genome, the individuals will eventually disappear because of genetic load. But if by chance a copy insert in a piRNA cluster, transposition rate will be controlled and individuals will survive. Examples can be found in the literature describing such events. *Drosophila virilis* is apparently in the process of ongoing *Penelope* reinvasion that is producing in some strains 27 nt piRNA in ovaries. Introduction of *Penelope* by transformation in the *D. melanogaster* genome conducted eventually in few years to the production of piRNA correlated with localization of *Penelope* copies in the 42 AB piRNA

cluster that are able to repress transposition (ROZHKOV *et al.* 2013). Another example, the *P* element naturally and successfully invaded the *Drosophila melanogaster* and *Drosophila simulans* genomes (ANXOLABEHERE *et al.* 1988; KOFLER *et al.* 2015). For the *P* element in *D. melanogaster* studied for 30 years, the precise mechanism of its auto-repression is partially understood. This involved that *P* elements have a hot spot of insertion within subtelomeric sequences of the *X* chromosome (*TAS-X*) that is a piRNA cluster. Thanks to the variety of *P* element derived transgenes, we are also starting to have a pretty clear global picture of the machinery involved in its regulation (JOSSE *et al.* 2007; TODESCHINI *et al.* 2010).

These two examples of genome dealing with new TE were studied while the system of repression has already reached a steady state. The question that we asked here was what is happening at the beginning of the process? In other words, do TE immediately produce piRNA because the piRNA cluster transcription will enclose the TE just inserted or alternatively does the TE sequence will have first to lose its properties before acquiring the piRNA cluster properties?

To address these questions, we made use of a known system based on *P[lArB]* *P*-transgenes inserted in *TAS-X* piRNA cluster. It has been shown that the *P[lArB]* transgene displays an epigenetic transmission of this repression capacities through meiosis, which involves maternally transmitted piRNA (JOSSE *et al.* 2007; TODESCHINI *et al.* 2010; DE VANSAY *et al.* 2012; MUERDTER *et al.* 2012). Then *P[lArB]* piRNA were found associated with the three PIWI proteins engaged in the ping-pong amplification loop and H3K9me3 marked were found on the telomeric transgene most likely initiated by piRNA associated Piwi that can recruit the RDC complex (LE THOMAS *et al.* 2014).

We first showed that *TAS-X* locus can be fully and strongly activated in only one generation in progenies that inherited the *TAS-X* locus from the father without maternal *TAS-X* piRNA. This efficient and complete activation might be interpreted as a kind of paramutation that involved piRNA produced from maternal autosomal piRNA clusters (*TAS-2R*, *TAS-3R* and 42AB) maintaining a constant level of piRNA production independently of its inheritance. Such paramutation appears difficult to formerly confirm because it will require the existence of a strain devoid of *TAS* at the 2R, 3R that we could not find (ASIF-LAIDIN *et al.* in preparation) as well as no *INV-4* in the genome that appear to be an ancient *D. melanogaster* element.

As a consequence, a transgene inserted in *TAS-X* locus can be considered to be a tool that mimics what is happening when a new TE inserts in a piRNA cluster. Several *scenarii* can be proposed. First a complete and strong repression is established immediately because the RDC complex is already functional on the locus and because maternally inherited piRNA loaded on Piwi are bombing the *TAS-X* locus surrounding the transgene. A transcript started in *TAS-X* and going through *P[lArB]* will be produced. Subsequently this

hybrid RNA will be then exported in the cytoplasm thanks to UAP56 and then enters in the ping-pong processing in the nuage. Alternatively, RDC complex might need several generations to settle on the transgene and erase the signals of a bona fide mRNA (splicing, termination of transcription, ...) present on the different regions of *P[lArB]* and replace to a bona fide piRNA precursors.

By analyzing maternally and paternally inherited *P[lArB]* transgene and *TAS-X* locus, we found that the transgenerational inheritance of *P[lArB]* piRNA is acquired progressively across generations (3 to 4) that is correlated to the acquisition of repression capacity. This progressive reinforcement of piRNA production appears to be independent of the presence of a target to repress. Surprisingly the efficacy of piRNA production from the *P[lArB]* transgene and the *TAS-X* locus do not follow the same kinetics. This might suggest that the RDC complex present is unable to process a new inserted sequence in a single generation. Therefore we try to identify the dynamic of RDC recruitment on the *P[lArB]* transgene by testing some of the properties associated with piRNA precursors. We found that inhibition of splicing and read through transcripts within *P[lArB]* transgene is occurring with the same kinetics, even if this two events are not occurring at 100%. We were also able to detect only few overlapping piRNA between *P[lArB]* and *TAS-X* suggesting that some piRNA precursors including *TAS-X* locus and transgenes might be synthesized.

Interestingly transgenerational inheritance observed in this study includes accumulation of piRNA across generations, when the *P[lArB]* is paternally inherited, daughters are stronger piRNA producers than their mothers, but conversely when the *P[lArB]* is maternally inherited starting from a homozygous locus to a heterozygous locus; piRNA production is decreasing, the mothers are stronger piRNA producers than their daughters. These two mirrored kinetics might reflect that the quantity of maternally transmitted piRNA as well as the chromosomal locus are required for the establishment of the equilibrium state that appears to be reached in four to five generations in both inheritance. Thus, the emergence of *P[lArB]*-PI is carrying out with the same rate than the *P[lArB]*-MI descent: epigenetic conversion can conduct to emergence or descent of a piRNA cluster up to an equilibrium state

Taken together, we are presenting a study about the establishment of a transgenerational epigenetic phenomenon that need more than one fly life and let an epigenetic imprint at each generation. This progressive competence to repress might have an impact on TE evolution when entering into a naive genome giving them opportunity to amplify during a short period before being completely repressed.

Acknowledgements

We thank the Bloomington Stock Center for providing stocks, FlyBase.org for providing valuable databases and ARTbio/Bioinformatics analyses platform at the IBPS for the development of valuable tools. We thank Ritha Zamy for technical assistance. This work was supported by fellowships from the Ministère de l'Enseignement Supérieur et de la Recherche to A.A-L and by grants from the Association de la Recherche contre le Cancer (Fondation ARC, SFI20121205921, SFI20131200470), the Fondation pour la Recherche Médicale, (FRM DEP20131128532), the Association Nationale de la Recherche (ANR, project "plastisipi") to S.R., and the University Pierre et Marie Curie (Emergence EME1223) to L.T.

Figure Legends

Figure. 1. X chromosome subtelomeric region of the *P-1152w^c* line

The *TAS-X* repeats are composed by solo LTR of *Inv4* (brown boxes), flanked by domains that can be found on *TAS-2R* and/or *TAS-3R* (grey boxes) and a specific 909 bp domain for *TAS-X*, called domain F (red boxes) (ASIF-LAIDIN *et al.* in preparation). Domain F will be used as the reference for the whole *TAS-X* region. The *P-1152w^c* line possesses 2 enhancer trap *P[lArB]* transgenes inserted in the subtelomeric repeats of the *X* chromosome (*TAS-X*). For simplification purposes, one transgene is presented in detail, the position of the second is symbolized by a star (JOSSE *et al.* 2008). The two transgenes are inserted in the same orientation and at the same nucleotide position within the solo LTR subrepeats of *INV-4*. Arrows represent the sense of transcription of the genes included in the transgene. The *lacZ* gene is expressed under the *P* element promoter, the "Hsp70Ab" contained the convergent 3' end of the *Hsp70Ab* and *CG3281* genes, "Adh" contains the upstream region of the *Adh/Adhr* region, missing the 2 last exons, "rosy" contains the upstream, downstream and coding region of the *rosy* gene, "pBluescript" was used to easily clone flanking sequences (WILSON *et al.* 1989) (for more details see **Fig. 5** and **Supplementary Fig. 7**). Note that sense RNA from the entire *TAS-X* locus including the *P[lArB]* transgene was arbitrary attributed to transcripts having a telomere towards centromere sense of transcription from later on.

Figure 2. Transgenerational inheritance of new piRNA production from a paternally acquired sequence inserted into *TAS-X*

Mating scheme used to produce maternal inheritance (*P[lArB]-MI*) (**A**) and paternal inheritance (*P[lArB]-PI*) (**B**) of the *P[lArB]* transgene. At each generation, five heterozygous females were crossed with five males carrying a *P-lacZ* target transgene (*BQ16*) to quantify the repression capacities (% of TSE) acquired through successive generations by performing an ovarian LacZ staining. The level of TSE was measured at each generation up to the ninth (see **Supplementary Table 2**). In parallel, for each lineage ovarian small RNA were extracted from 20 females from the first to the fourth generation for maternal (subline H) and paternal (subline D) inheritance and from the first to the third generation for subline B. (**C**) The graph shows TSE repression for the first four generations (means for the four sets of sublines and standard deviation) of the two types of mating: *P[lArB]-MI* in red and *P[lArB]-PI* in blue. The capacity of repression (TSE) for the paternal inheritance *P[lArB]-PI* increases progressively and acquires full level of repression after three to four generations. *P[lArB]-MI* presents a constant full repression. (**D**) Normalized piRNA distribution on subline D plotted onto the domain F sequence shows that piRNA are homogeneously distributed through

the sequence and is constant over generations. On the contrary, *P[lArB]* transgene piRNA in subline D distributed homogeneously along the sequence are increasing in quantity over the generations (F). The structure of *P[lArB]* transgene is represented under the graph. (E and G) Size distribution for the piRNA of subline D corresponding to the domain F (E) and *P[lArB]* (G). The two sequences harbor a different profile. The domain F shows a constant size distribution over generations, while the transgene shows a huge amount of siRNA compared to the piRNA at the first generation that ratio is decreasing over generation (see Fig.4C). (H) The Graph represents the ratio of paternal (subline D) over maternal (subline H) normalized piRNA reads for the *P[lArB]* transgene as well as the different sequences (*lacZ*, *Hsp70*, *Adh*, *rosy* and *pBluescript*) constituting the transgene as well as for the domain F. While the domain F, representing the entire *TAS-X* piRNA cluster, produces a constant quantity of piRNA, all the sequences have the same profile, that is to say on G1 the ratio is low and then increase progressively over generations. At the fourth generation, the ratio of *P[lArB]* transgene reaches the ratio of domain F. The system archives the equilibrium at four to five generations.

Figure 3. Transgenerational inheritance of *P[lArB]* and *TAS-X* piRNA

(A) Normalized domain F piRNA reads and (B) Normalized *P[lArB]* piRNA reads from the maternal and paternal inheritance were used to construct the graphs (Supplementary Table 3). Homozygous *P-1152* for the *P[lArB]* transgene was added as G0 for the maternal inheritance, and *Canton^v* devoid of transgene was added as G0 for the paternal inheritance (Supplementary Table 4). (A) Histogram shows that homozygous *P[lArB]* strain produces about two times more domain F piRNA than the heterozygous. Regardless to the inheritance, from the first generation, the same amount of piRNA are synthesized from the domain F. (B) On the contrary, while homozygous *P[lArB]* is losing progressively his capacity to produce piRNA when it is heterozygously transmitted, the paternally inherited *P[lArB]* locus is acquiring progressively the capacity to produce piRNA. For both crosses, the equilibrium state seems to be reached in about five generations.

Figure 4. *P[lArB]* property shift in the *P[lArB]*-MI and *P[lArB]*-PI lineages across generations

P-1152^{w^c} line that carries the homozygous *P[lArB]* transgene was included as a representative of G0 for the maternal lineage and *Canton^v* strain devoid of transgene was also considered as a G0 for the paternal lineage. (A) The ratios of normalized reads per kb for exons vs introns were calculated for the *Adh* and *rosy* sequences included in the *P[lArB]* transgene (Supplementary Table 5). The *rosy* gene that is not or weakly expressed in ovaries displays a constant ratio over the generations no matter the inheritance. On the contrary,

the *Adh* gene that is expressed shows a different profile. For the *P[lArB]-MI*, *P[lArB]* has a piRNA transcription where the ration is constant over the generations, whereas for the *P[lArB]-PI*, *P[lArB]* has in G1 an mRNA pattern of transcription with more than thirteen fold more exons than introns. This ratio is dropping from the second generation to reach a ratio close to maternal ratio and even the ratio observed for non-expressed genes. (B) The ratio of normalized reads between G1 and G4 for sense vs antisense normalized reads (Supplementary Table 6). (C) The ratio of Paternal and maternal siRNA (21nt) versus piRNA (23-29nt) normalized reads for the *P[lArB]* transgene shows that the amount of siRNA in the G1 of *P[lArB]-PI* dropped from the G2 (Supplementary Table 7). We hypothesized that at the beginning of the process double stranded RNA are synthesized but the piRNA machinery is not fully functional giving a substrate to the siRNA machinery. Then the RDC complex started to be efficient and piRNA are produced.

Figure 5. piRNA production from *Adh* and *rosy* genes included in the *P[lArB]* transgene

The raw values of 23 to 29nt small RNA have been plotted against *D. melanogaster* genome (R5.9) under IGV (Integrative Genomes Viewer) in 6 small RNA libraries *Canton^s*, paternal subline D from generation 1 to 4 and maternal subline H generation. (A) *Adh* and *rosy* regions present in the *P[lArB]* are mark out in red. The *Adh* region contains the upstream promoting regions and lacks the two last exons. Only few isoforms of *Adh* are not represented. (B) The *rosy* region in the *P[lArB]* contains upstream and downstream flanking sequence of *rosy* gene. The upstream sequence includes a part of the *l(3)87Df* gene. The black stars represent the overlapping small RNA between exon and intron.

Supplementary Fig. 1. Mapping of maternally inherited piRNA on *P[lArB]*, subline H

Distribution of 23-29 nt piRNA isolated from the subline H. piRNA have been mapped on the sequence of domain F, a subdomain of *TAS-X*, on the full length of the *P[lArB]* transgene, and on the different parts of the transgene, *i.e.* *lacZ* gene, *Hsp70*, *Adh*, *rosy*, *pBluescript* (see Fig. 1). The sense piRNA in red are arbitrary defined as the piRNA produced by the telomere towards centromere strand, the antisense piRNA in blue are from the reverse strand.

Supplementary Fig. 2. Mapping of paternally inherited piRNA on *P[lArB]*, subline D

Distribution of 23-29 nt piRNA isolated from the subline D. piRNA have been mapped on the sequence of domain F, a subdomain of *TAS-X*, on the full length of the *P[lArB]* transgene, and on the different parts of the transgene, *i.e.* *lacZ* gene, *Hsp70*, *Adh*, *rosy*, *pBluescript* (see Fig. 1). The sense piRNA in red are arbitrary define as the piRNA produced by the telomere towards centromere strand, the antisense piRNA in blue are from the reverse strand. NB: Distribution of piRNA on the domain F and the *P[lArB]* transgene is reproduced from **Fig. 2D** and **2F**.

Supplementary Fig. 3. Mapping of paternally inherited piRNA on *P[lArB]*, subline B

Distribution of 23-29 nt piRNA isolated from the subline B. piRNA have been mapped on the sequence of domain F, a subdomain of *TAS-X*, on the full length of the *P[lArB]* transgene, and on the different part of the transgene, *i.e.* *lacZ* gene, *Hsp70*, *Adh*, *rosy*, *pBluescript* (see **Fig. 1**). The sense piRNA in red are arbitrary defined as the piRNA produced by the telomere towards centromere strand, the antisense piRNA in blue are from the reverse strand.

Supplementary Fig. 4. Size distribution of small RNA in the *P[lArB]-MI*, subline H

Size distribution of small RNA isolated from the *P[lArB]-MI* subline H matching to the sequences included in the *P[lArB]* transgene and in the Domain F. The sense piRNA in red are arbitrary defined as the piRNA produced by the telomere towards centromere strand, the antisense piRNA in blue are from the reverse strand.

Supplementary Fig. 5. Size distribution of small RNA in the *P[lArB]-PI*, subline D

Size distribution of small RNA isolated from the *P[lArB]-PI* subline D matching to the sequences included in the *P[lArB]* transgene and in the Domain F. The sense piRNA in red are arbitrary defined as the piRNA produced by the telomere towards centromere strand, the antisense piRNA in blue are from the reverse strand. NB: Size distribution of small RNA on the domain F and the *P[lArB]* transgene is reproduced from **Fig. 2E** and **2G**.

Supplementary Fig. 6. Size distribution of small RNA in the *P[lArB]-PI*, subline B

Size distribution of small RNA isolated from the *P[lArB]-PI* subline B matching to the sequences included in the *P[lArB]* transgene and in the Domain F. The sense piRNA in red are arbitrary defined as the piRNA produced by the telomere towards centromere strand, the antisense piRNA in blue are from the reverse strand.

Supplementary Fig. 7. piRNA production from *Hsp70* genomic region included in the *P[lArB]* transgene

The raw values of 23 to 29nt small RNA have been plotted under IGV (Integrative Genomes Viewer) against *D. melanogaster* genome (R5.9) in 6 small RNA libraries *Canton^s*, paternal subline D from G1 to G4 and maternal subline H generation. The *Hsp70* region in the *P[lArB]* (mark out in red) is composed by two transcription terminators, one from the *Hsp70Ab* and one from *CG3281* and also includes about 600bp of *CG3281* exon. Blue and red reads the anti-sense and sense reads respectively. Grey reads correspond to reads mapping on multiple sites, here on the duplicated *Hsp70Ab* and *Hsp70Aa* genes.

References

- Anxolabehere, D., M. G. Kidwell and G. Periquet, 1988 Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile P elements. *Mol Biol Evol* 5: 252-269.
- Asif-Laidin, A., V. Delmarre, J. Laurentie, W. Miller, S. Ronsseray *et al.*, in preparation Telomeric Associated Sequences in *Drosophila*: structure, dynamic and properties of a piRNA cluster.
- Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis *et al.*, 2007 Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128: 1089-1103.
- Brennecke, J., C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark *et al.*, 2008 An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322: 1387-1392.
- Czech, B., and G. J. Hannon, 2016 One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem Sci*.
- de Vanssay, A., A. L. Bouge, A. Boivin, C. Hermant, L. Teyssset *et al.*, 2012 Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature* 490: 112-115.
- Dufourt, J., C. Dennis, A. Boivin, N. Gueguen, E. Theron *et al.*, 2014 Spatio-temporal requirements for transposable element piRNA-mediated silencing during *Drosophila* oogenesis. *Nucleic Acids Res* 42: 2512-2524.
- Fanti, L., D. R. Dorer, M. Berloco, S. Henikoff and S. Pimpinelli, 1998 Heterochromatin protein 1 binds transgene arrays. *Chromosoma* 107: 286-292.

- Fransz, P. F., S. Armstrong, J. H. de Jong, L. D. Parnell, C. van Drunen *et al.*, 2000 Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell* 100: 367-376.
- Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura *et al.*, 2007 A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315: 1587-1590.
- Han, B. W., W. Wang, C. Li, Z. Weng and P. D. Zamore, 2015 Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* 348: 817-821.
- Hermant, C., A. Boivin, L. Teyssset, V. Delmarre, A. Asif-Laidin *et al.*, 2015 Paramutation in *Drosophila* Requires Both Nuclear and Cytoplasmic Actors of the piRNA Pathway and Induces Cis-spreading of piRNA Production. *Genetics* 201: 1381-1396.
- Hirakata, S., and M. C. Siomi, 2016 piRNA biogenesis in the germline: From transcription of piRNA genomic sources to piRNA maturation. *Biochim Biophys Acta* 1859: 82-92.
- Iwasaki, Y. W., M. C. Siomi and H. Siomi, 2015 PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem* 84: 405-433.
- Josse, T., C. Maurel-Zaffran, A. de Vanssay, L. Teyssset, A. L. Todeschini *et al.*, 2008 Telomeric trans-silencing in *Drosophila melanogaster*: tissue specificity, development and functional interactions between non-homologous telomeres. *PLoS One* 3: e3249.
- Josse, T., L. Teyssset, A. L. Todeschini, C. M. Sidor, D. Anxolabehere *et al.*, 2007 Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS Genet* 3: 1633-1643.
- Karpen, G. H., and A. C. Spradling, 1992 Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics* 132: 737-753.
- Klattenhoff, C., H. Xi, C. Li, S. Lee, J. Xu *et al.*, 2009 The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* 138: 1137-1149.
- Kofler, R., T. Hill, V. Nolte, A. J. Betancourt and C. Schlotterer, 2015 The recent invasion of natural *Drosophila simulans* populations by the P-element. *Proc Natl Acad Sci U S A* 112: 6659-6663.
- Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25.
- Le Thomas, A., E. Stuwe, S. Li, J. Du, G. Marinov *et al.*, 2014 Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev* 28: 1667-1680.
- Mohn, F., D. Handler and J. Brennecke, 2015 Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* 348: 812-817.
- Mohn, F., G. Sienski, D. Handler and J. Brennecke, 2014 The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* 157: 1364-1379.
- Muerdter, F., I. Olovnikov, A. Molaro, N. V. Rozhkov, B. Czech *et al.*, 2012 Production of artificial piRNAs in flies and mice. *RNA* 18: 42-52.
- O'Kane, C. J., and W. J. Gehring, 1987 Detection in situ of genomic regulatory elements in *Drosophila*. *Proc Natl Acad Sci U S A* 84: 9123-9127.
- Pimpinelli, S., M. Berloco, L. Fanti, P. Dimitri, S. Bonaccorsi *et al.*, 1995 Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. *Proc Natl Acad Sci U S A* 92: 3804-3808.
- Poyhonen, M., A. de Vanssay, V. Delmarre, C. Hermant, A. L. Todeschini *et al.*, 2012 Homology-dependent silencing by an exogenous sequence in the *Drosophila* germline. *G3 (Bethesda)* 2: 331-338.
- Roche, S. E., and D. C. Rio, 1998 Trans-silencing by P elements inserted in subtelomeric heterochromatin involves the *Drosophila* Polycomb group gene, Enhancer of zeste. *Genetics* 149: 1839-1855.

- Ronsseray, S., M. Lehmann and D. Anxolabehere, 1989 Copy number and distribution of P and I mobile elements in *Drosophila melanogaster* populations. *Chromosoma* 98: 207-214.
- Ronsseray, S., L. Marin, M. Lehmann and D. Anxolabehere, 1998 Repression of hybrid dysgenesis in *Drosophila melanogaster* by combinations of telomeric P-element reporters and naturally occurring P elements. *Genetics* 149: 1857-1866.
- Rozhkov, N. V., N. G. Schostak, E. S. Zelentsova, I. A. Yushenova, O. G. Zatsepina *et al.*, 2013 Evolution and dynamics of small RNA response to a retroelement invasion in *Drosophila*. *Mol Biol Evol* 30: 397-408.
- Shpiz, S., S. Ryazansky, I. Olovnikov, Y. Abramov and A. Kalmykova, 2014 Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the *Drosophila* germline. *PLoS Genet* 10: e1004138.
- Theron, E., C. Dennis, E. Brasset and C. Vauray, 2014 Distinct features of the piRNA pathway in somatic and germ cells: from piRNA cluster transcription to piRNA processing and amplification. *Mob DNA* 5: 28.
- Todeschini, A. L., L. Teyssset, V. Delmarre and S. Ronsseray, 2010 The epigenetic trans-silencing effect in *Drosophila* involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1. *PLoS One* 5: e11032.
- Wang, W., B. W. Han, C. Tipping, D. T. Ge, Z. Zhang *et al.*, 2015 Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. *Mol Cell* 59: 819-830.
- Wilson, C., R. K. Pearson, H. J. Bellen, C. J. O'Kane, U. Grossniklaus *et al.*, 1989 P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. *Genes Dev* 3: 1301-1313.
- Zhang, Z., J. Wang, N. Schultz, F. Zhang, S. S. Parhad *et al.*, 2014 The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* 157: 1353-1363.

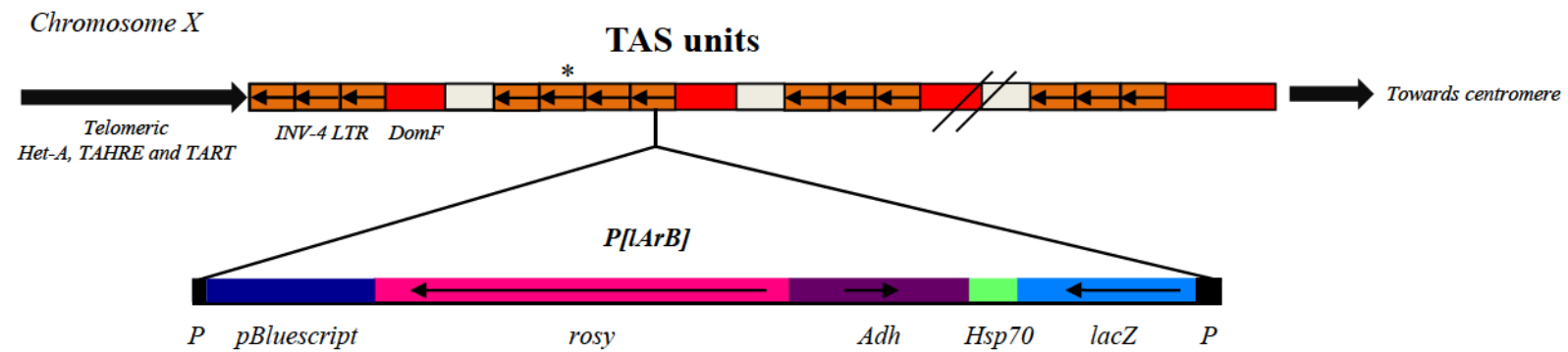
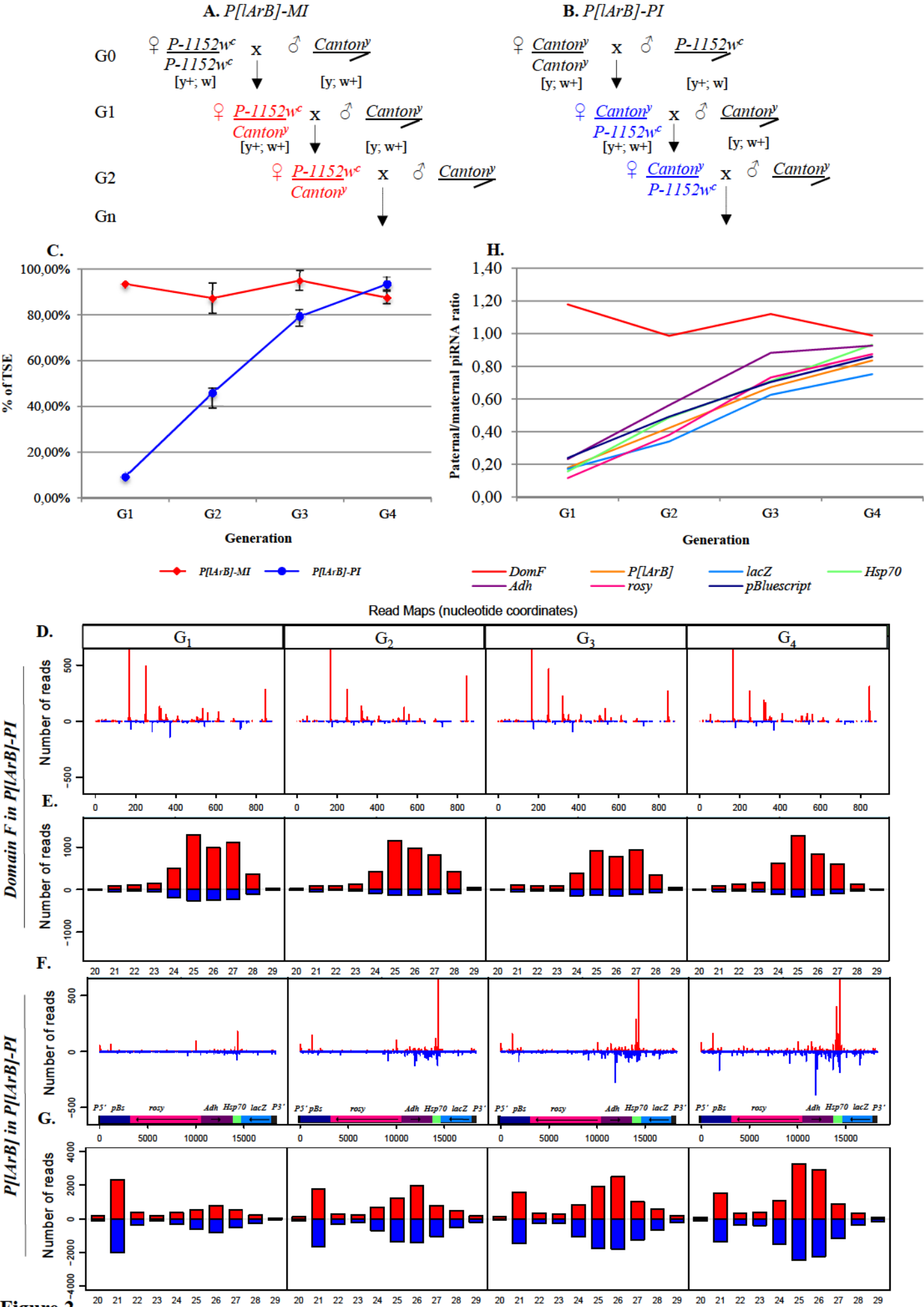


Figure 1



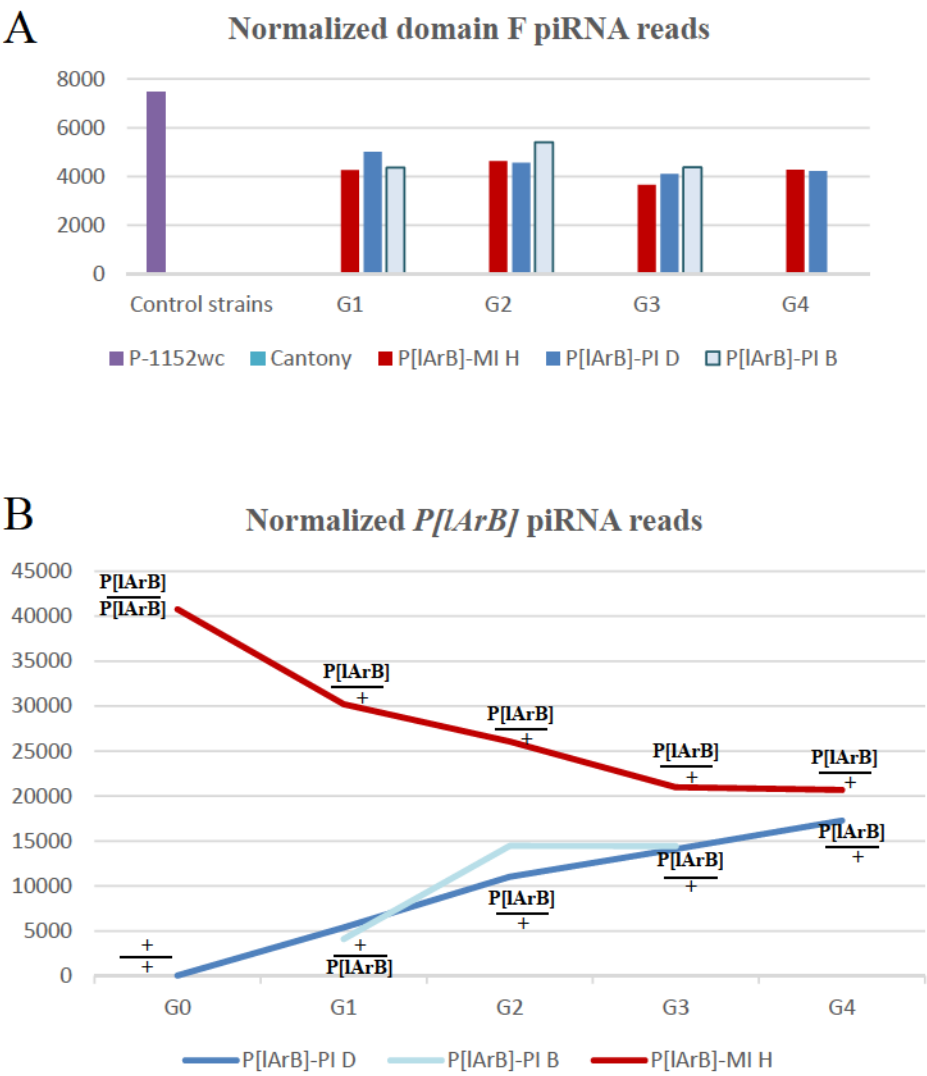
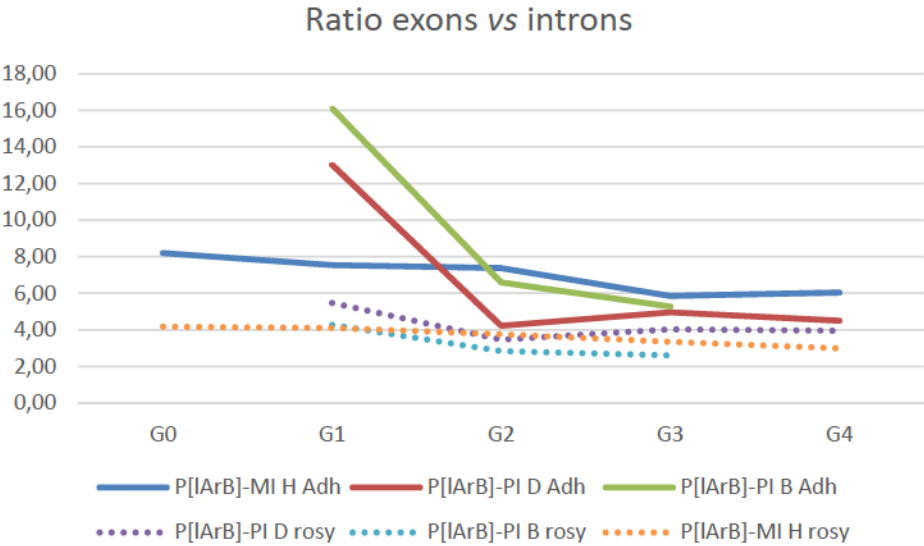
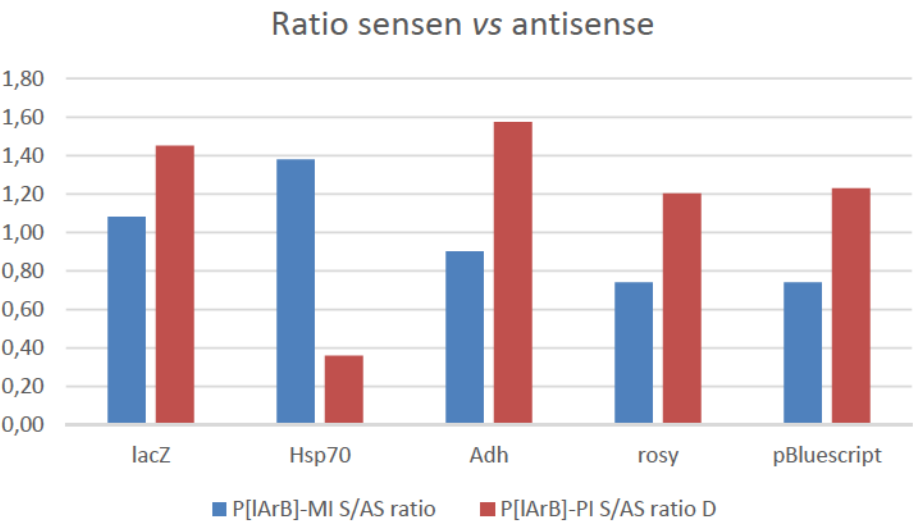


Figure 3

A



B



C

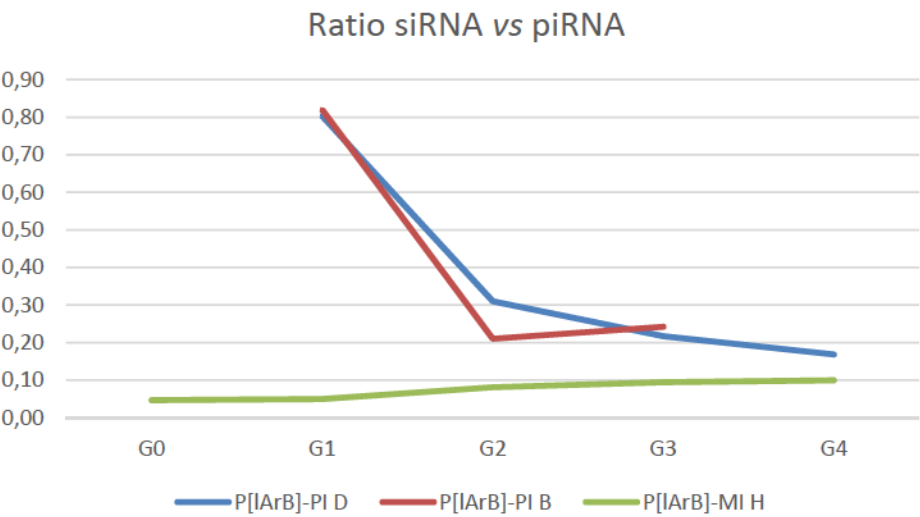


Figure 4

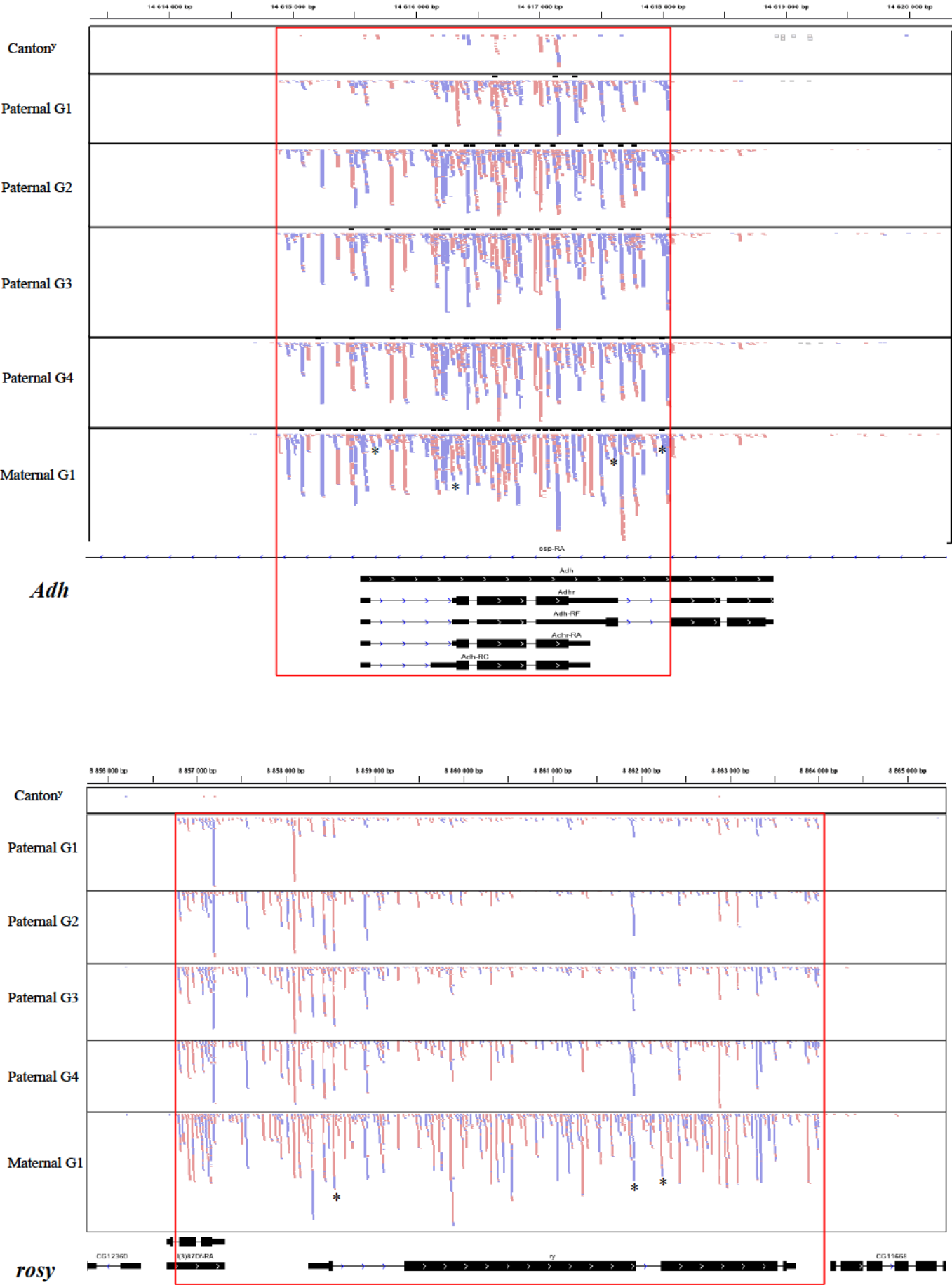
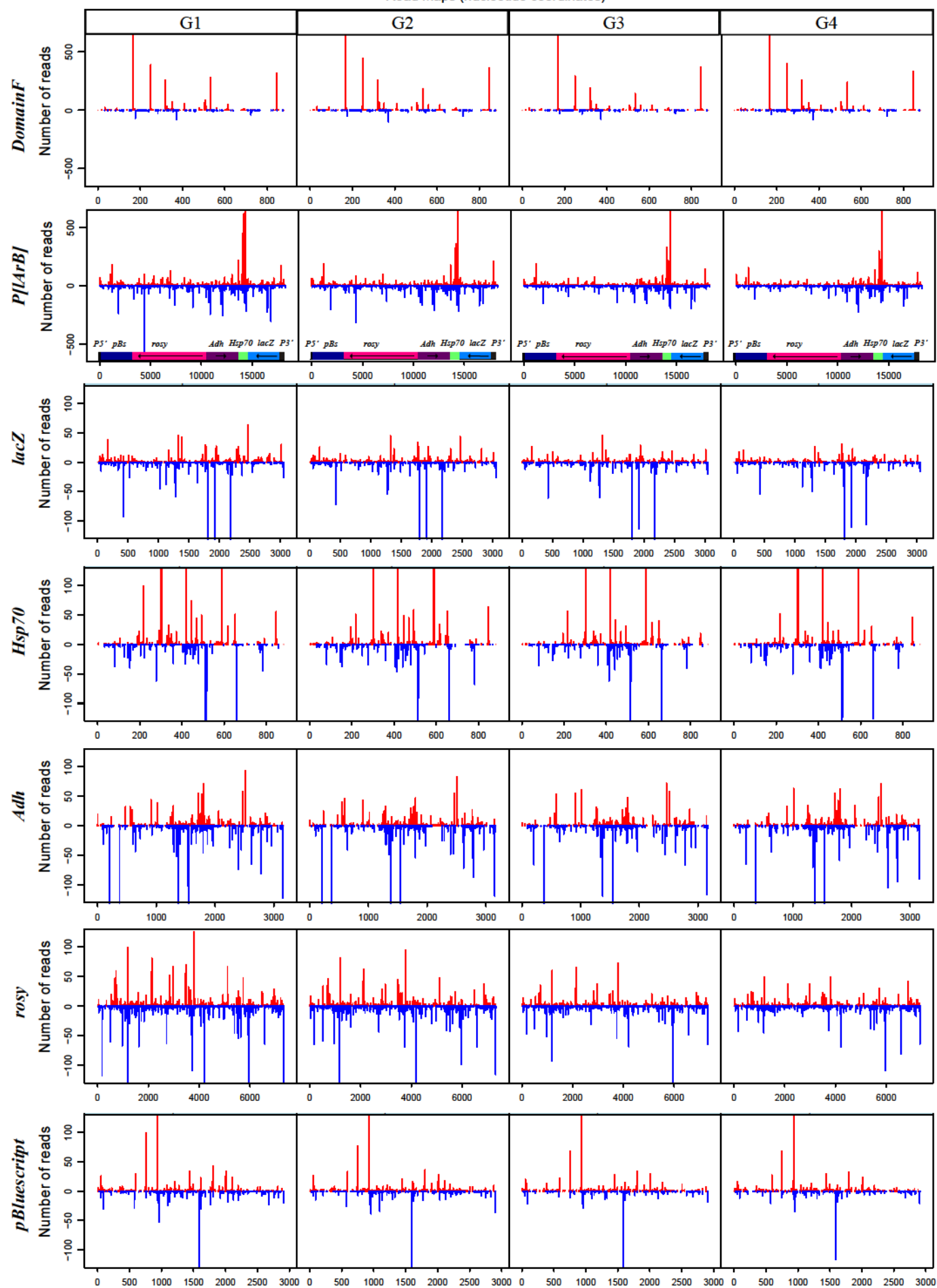


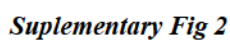
Figure 5

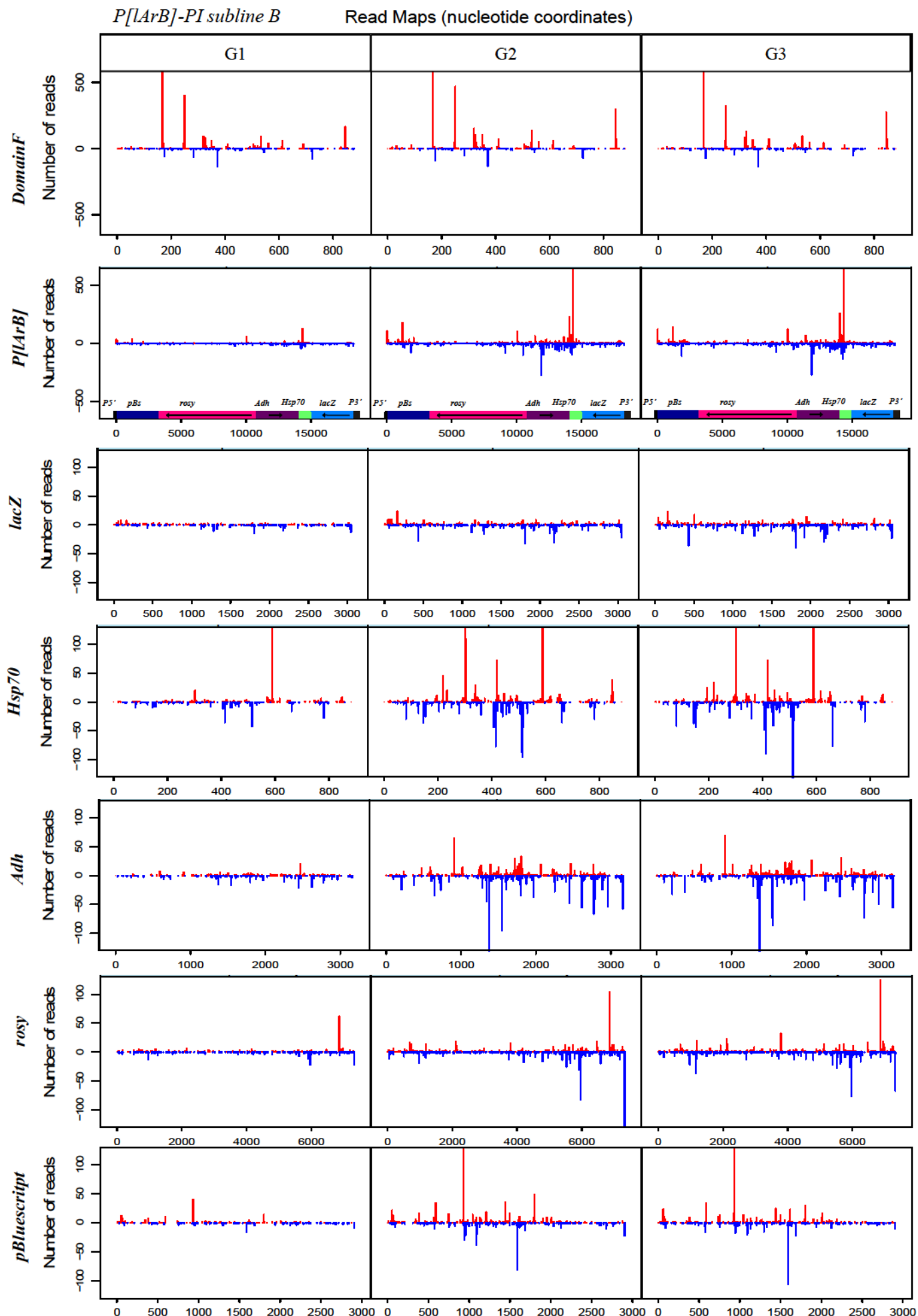
P[lArB]-MI subline H

Read Maps (nucleotide coordinates)

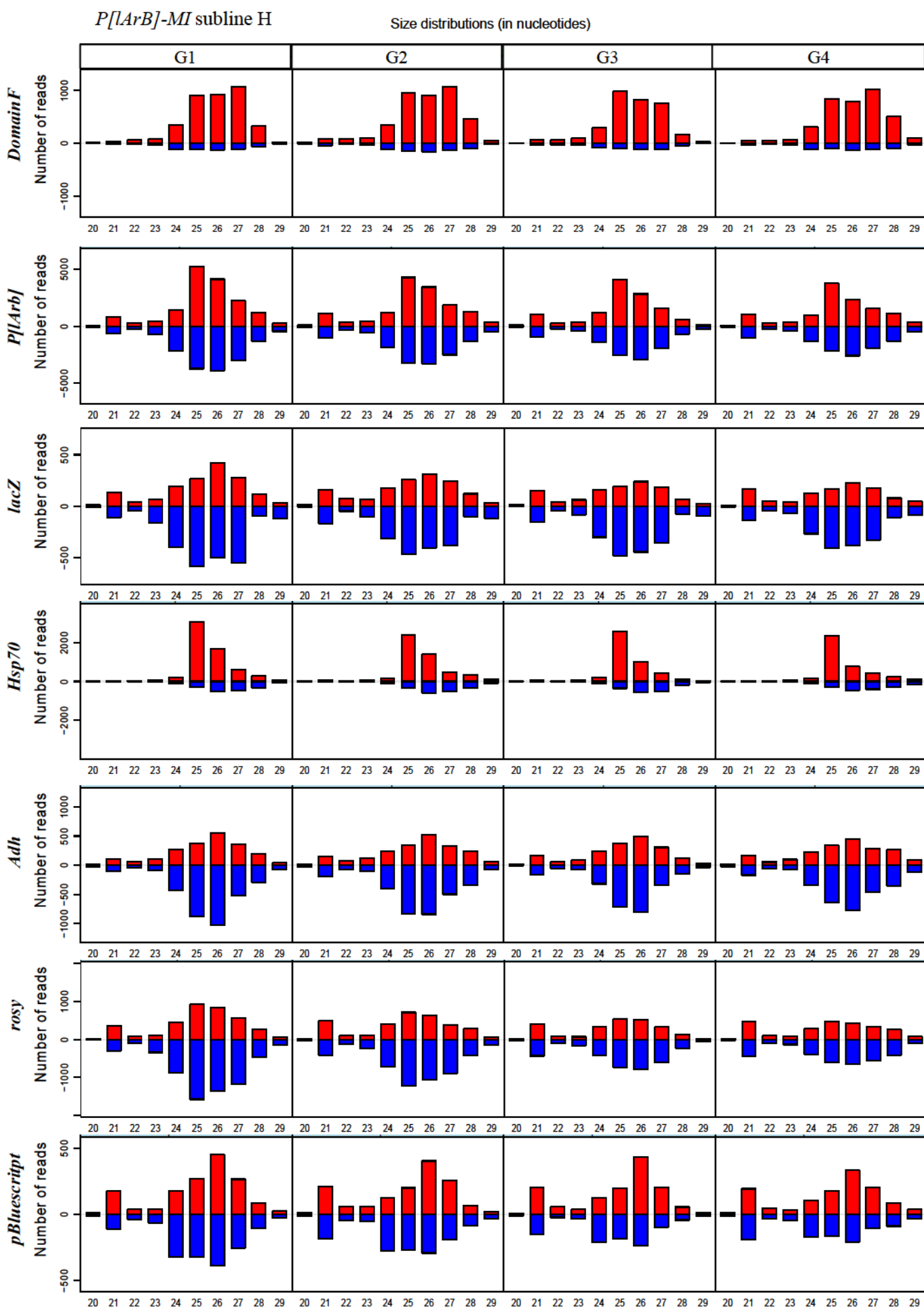


Supplementary Fig 1

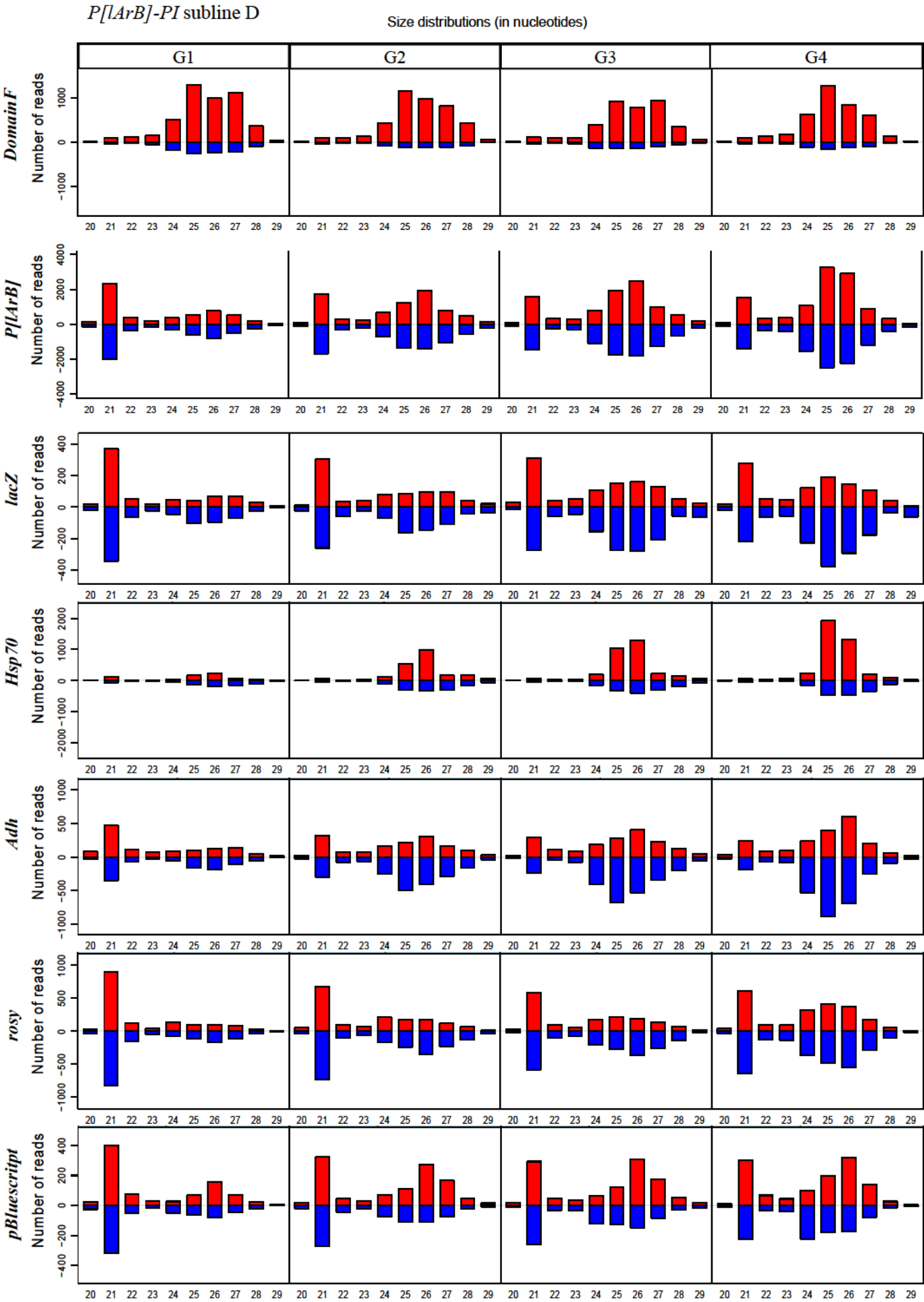




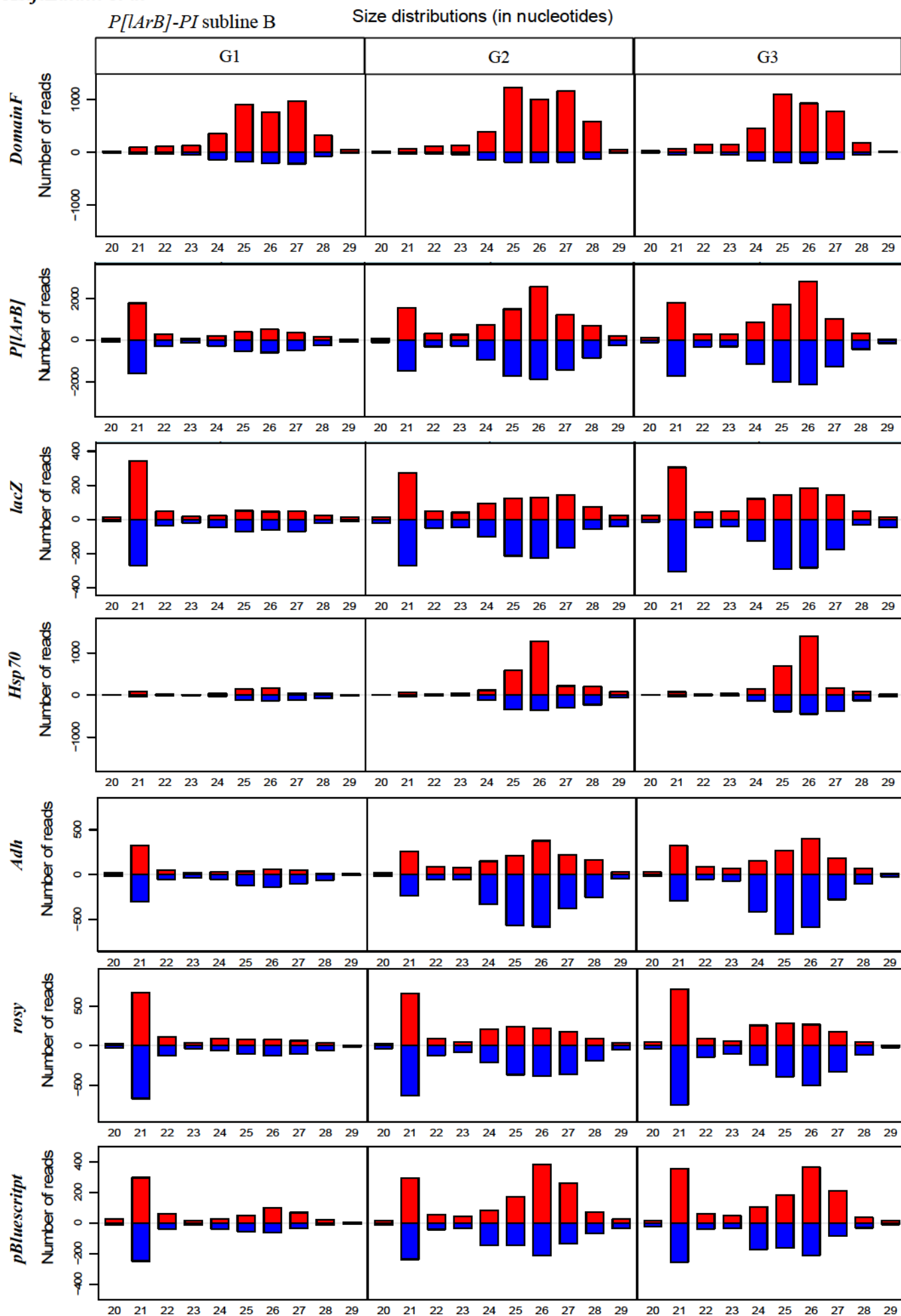
Supplementary Fig. 3



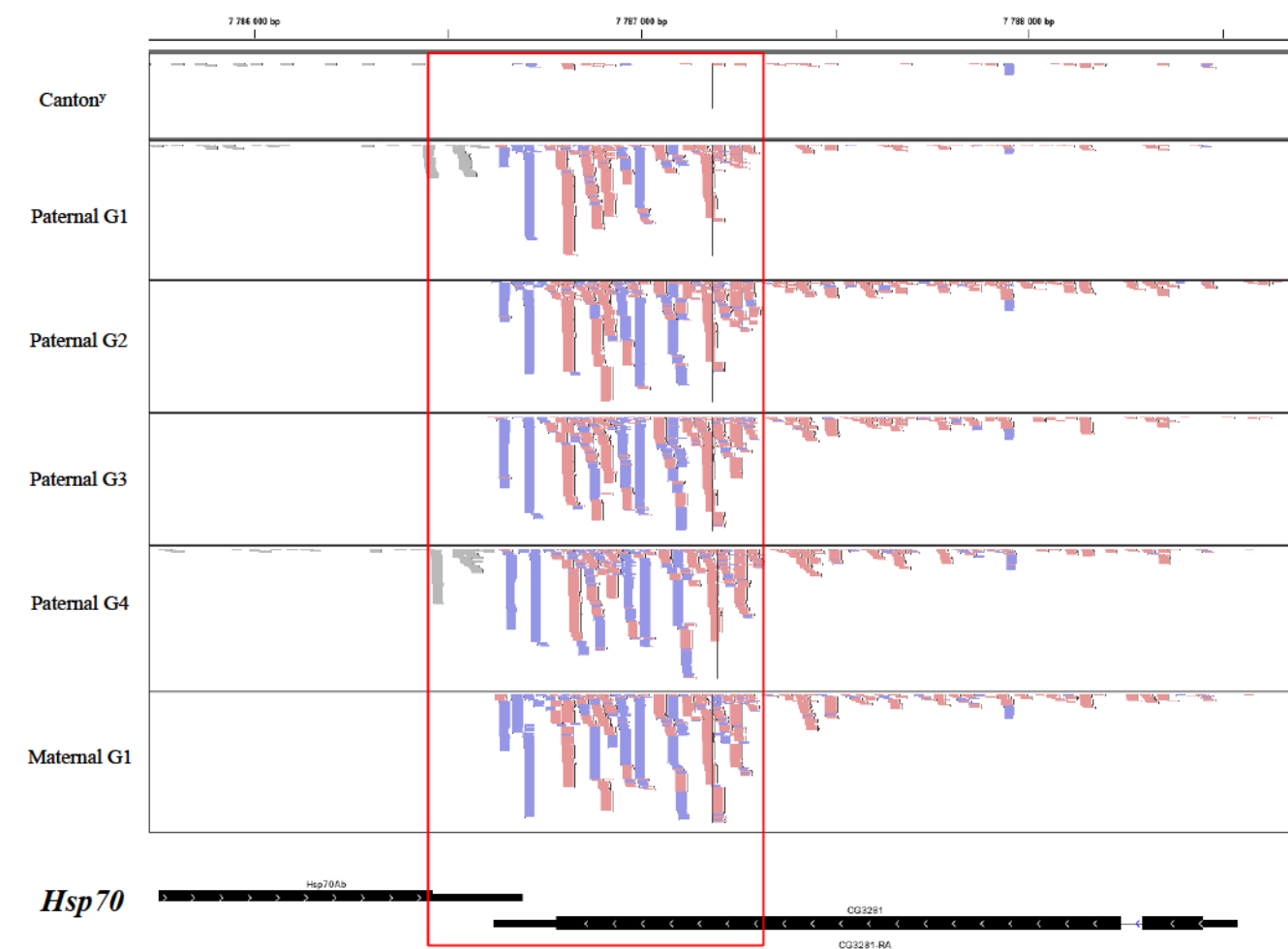
Supplementary Fig. 4



Supplementary Fig. 5



Supplementary Fig. 6



Supplementary Table 1 : Annotation of small RNA libraries

Library	1	2	3	4	5	6
Subline and generations	Maternal H G1	Maternal H G2	Maternal H G3	Maternal H G4	Paternal B G1	Paternal D G1
alignment reference	GRH117	GRH118	GRH119	GRH120	GRH121	GRH122
D. melanogaster (Flybase r5.49)	8749771	8010061	7614649	7092541	8792420	8365118
Dmel_all-miscRNA	401121	964415	502779	542585	5455	839017
Dmel_all-tRNA	122722	215323	108824	118454	63501	138242
Dmel_all-transposon	4 206 089	3 362 562	3 447 748	3 219 305	4 058 452	3 707 867
Dmel_miR_r20	1 959 665	1 714 231	1 803 742	1 546 512	1 998 328	1 732 825
Remaining Unmatched	2 060 174	1 753 530	1 751 556	1 665 685	2 126 639	1 947 167
effective depth	8225928	6830323	7003046	6431502	8723464	7387859
Facteur Normalisation	0,78	0,94	0,92	1,00	0,74	0,87

Library	7	8	9	10	11
Subline and generations	Paternal B G2	Paternal D G2	Paternal B G3	Paternal D G3	Paternal D G4
alignment reference	GRH123	GRH124	GRH126	GRH127	GRH128
D. melanogaster (Flybase r5.49)	9448234	9092202	9421670	7925059	7751589
Dmel_all-miscRNA	550691	764979	908722	55603	454261
Dmel_all-tRNA	192744	20465	212698	117016	99062
Dmel_all-transposon	4 502 347	4 084 503	4 100 466	3 537 130	3 417 567
Dmel_miR_r20	1 879 526	1 965 293	2 049 340	1 885 826	2 119 232
Remaining Unmatched	2 322 926	2 072 777	2 150 444	1 829 057	1 661 467
effective depth	8704799	8306758	8300250	7752440	7198266
Facteur Normalisation	0,74	0,77	0,77	0,83	0,83

Small RNAs were prepared from ovaries of females of the indicated strains. Values for the different categories of sequences matching to the *D. melanogaster* genome 5.49 in each library are indicated. For comparisons, libraries were normalized (normalization factor) relatively to the number of sequence reads aligning to the *D. melanogaster* genome but not to tRNAs or miscRNAs (effective depth). Library #4 has the lowest effective depth and was taken as reference to normalize the other libraries.

Supplementary Table2. Measure of ovarian *lacZ* repression for the maternally and paternally inherited *P[lArB]* transgene

Sublines	Paternally transmitted <i>P[lArB]</i> silencer transgene: <i>P[lArB]-PI</i>				Maternally transmitted <i>P[lArB]</i> silencer transgene: <i>P[lArB]-PI</i>			
	A	B	C	D	E	F	G	H
G1	9,25% n=1000				93,5% n=1000			
G2	55,7% n=1100	44,35% n=1100	42,5% n=1350	41,2% n=750	89,7% n=1100	85,1% n=1350	86,75% n=1450	87,6% n=1550
G3	82,25% n=1300	83,75% n=1500	75,1% n=1400	76,25% n=1450	96,75% n=1350	93,55% n=1400	91,8% n=1500	98% n=1400
G4	96,3% n=1400	90,5% n=1350	91,95% n=1300	95,1% n=1600	91,35% n=1800	88,1% n=1650	84,15% n=1700	86,15% n=1350
G5	95,93% n=1350	NT	89,6% n=600	88,35% n=1300	92,4% n=1550	85,55% n=1250	NT	90,9% n=1250
G6	94,5% n=850	92,1% n=600	92,7% n=1050	98% n=1000	91,8% n=850	91,18% n=800	96,9% n=1000	91,85% n=650
G7	77,4% n=1200	98,75% n=1200	NA	NA	89,5% n=1100	80% n=2000	NA	71% n=1550
G8	13,5% n=1450	73,5% n=1000	NA	NA	78,1% n=1350	69,75% n=800	NA	73,1% n=1450
G9	46,4% n=700	87,4% n=700	NA	NA	96,5% n=800	79,5% n=400	NA	

To measure the level of *lac Z* repression, females of each crosses shown in Fig. 2A and B were crossed with males encoding an euchromatic *P-lacZ* transgene expressed in female germline. The germline repression (% of TSE) was scored by calculating the % of repressed egg chambers. Twelve females were tested for each cross. n is he number of egg chambers counted.

NT: Not Tested

NA: Not Applicable, when the subline was unfertile probably due to isogenization.

Supplementary Table 3.

A. Normalized 23 to 29 nt reads found in the 3 sublines sequenced

<i>P[lArB]-PI D</i>	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	5394,0	5012,0	649,0	1233,7	1208,4	1057,9	667,3
G2	11035,6	4562,3	1051,1	3391,9	2769,7	2769,7	1148,8
G3	14109,2	4098,5	1733,9	4447,1	3605,5	3605,5	1319,7
G4	17277,6	4223,9	1888,6	5498,4	4183,9	4183,9	1543,3

<i>P[lArB]-PI B</i>	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	4117,4	4360,1	526,1	990,9	738,5	882,1	510,6
G2	14475,1	5405,7	1481,5	3955,3	3445,4	2666,2	1812,3
G3	14433,7	4376,7	1691,7	4082,5	3284,1	2795,1	1663,2

<i>P[lArB]-MI H</i>	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	30211,7	4252,6	3774,4	7861,6	5204,2	9133,8	2786,9
G2	26079,4	4624,8	3091,7	6969,2	4926,5	7284,1	2335,0
G3	20991,6	3658,8	2771,0	6275,3	4085,7	4928,4	1874,0
G4	20680,0	4275,0	2512,0	5900,0	4516,0	4782,0	1796,0

B. Paternal (P) vs Maternal (M) ratio of normalized 23 to 29 nt reads

Ratio P/M 23-29nt D	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	0,18	1,18	0,17	0,16	0,23	0,12	0,24
G2	0,42	0,99	0,34	0,49	0,56	0,38	0,49
G3	0,67	1,12	0,63	0,71	0,88	0,73	0,70
G4	0,84	0,99	0,75	0,93	0,93	0,87	0,86

Ratio P/M 23-29nt B	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	0,14	1,03	0,14	0,13	0,14	0,10	0,18
G2	0,56	1,17	0,48	0,57	0,70	0,37	0,78
G3	0,69	1,20	0,61	0,65	0,80	0,57	0,89

Supplementary Table 4. Normalized 23-29nt reads for *Canton^y* and *P-1152w^c* strains.

Strain	<i>P[LArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
<i>Canton^y</i>	60	-	-	24	31	4	1
<i>P-1152w^c</i>	40769	7477	4773	8090	7825	13807	4316

Canton^y strains do not bear the *P[LArB]* transgene neither *TAS-X* sequences (Asif-laidin et al) and were used to test the capacity to an empty genome to produce piRNA corresponding to the sequences present in the *P[LArB]* transgene.

The *P-1152w^c* strain is homozygous for the *P[LArB]* transgene.

Supplementary Table 5. Intronic and Exonic piRNA reads included in the *P[lArB]* transgene

<i>P[lArB]-PI D Adh</i>	G1	G2	G3	G4
introns	43	246	273	343
exons	560	1040	1357	1544
Ratio exon/intron	13,03	4,22	4,96	4,50

<i>P[lArB]-PI B Adh</i>	G1	G2	G3
introns	21	208	244
exons	345	1370	1288
Ratio exon/intron	16,12	6,60	5,28

<i>P[lArB]-MI H Adh</i>	G1	G2	G3	G4
introns	294	282	293	307
exons	2220	2081	1717	1855
Ratio exon/intron	7,54	7,38	5,86	6,04

<i>P-1152w^c</i>	<i>Adh</i>
introns	505
exons	4143
Ratio exon/intron	8,20

<i>P[lArB]-PI D rosy</i>	G1	G2	G3	G4
introns	30	118	133	157
exons	166	410	536	622
Ratio exon/intron	5,48	3,47	4,04	3,95

<i>P[lArB]-PI B rosy</i>	G1	G2	G3
introns	31	131	149
exons	134	373	390
Ratio exon/intron	4,28	2,85	2,61

<i>P[lArB]-MI H rosy</i>	G1	G2	G3	G4
introns	340	293	217	232
exons	1400	1100	726	695
Ratio exon/intron	4,11	3,76	3,35	2,99

<i>P-1152w^c</i>	<i>rosy</i>
introns	578
exons	2415
Ratio exon/intron	4,18

Intronic and exonic reads correspond to the transcribe strand of each genes. Only sequences included in the *P[lArB]* were matching. In order to minimize the bias introduced by difference sizes between introns and exons, the reads were normalized per kb

Supplementary Table 6. Sense versus antisense ratio for the 23-29nt normalized piRNA of each region included in the *P[lArB]* transgene

<i>P[lArB]</i> -PI S/AS ratio D	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	0,98	4,09	0,74	0,84	0,96	0,83	1,34
G2	1,01	6,65	0,76	1,57	0,62	0,69	1,63
G3	1,01	5,31	0,60	1,96	0,57	0,62	1,35
G4	1,02	5,98	0,51	2,33	0,61	0,69	1,09

<i>P[lArB]</i> -PI S/AS ratio B	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	0,79	3,97	0,71	0,85	0,41	0,72	1,26
G2	0,99	5,08	0,73	1,73	0,56	0,61	1,34
G3	0,96	4,60	0,70	1,65	0,54	0,63	1,36

<i>P[lArB]</i> -MI S/AS ratio H	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	0,97	6,61	0,57	3,25	0,57	0,88	0,88
G2	0,97	5,65	0,64	2,47	0,60	0,93	0,93
G3	1,05	6,52	0,51	2,44	0,68	1,29	1,29
G4	1,02	5,99	0,53	2,36	0,64	1,18	1,18

P1152wc S/AS ratio	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
	0,93	6,58	0,56	2,70	0,65	0,66	0,92

Supplementary Table 7. Paternal and maternal siRNA (21nt) versus piRNA (23-29nt) normalized reads for each reference sequence

<i>P[lArB]-PI</i> siRNA/piRNA ratio D	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	0,80	0,03	1,09	0,16	0,68	1,64	1,07
G2	0,31	0,03	0,53	0,04	0,22	0,51	0,52
G3	0,22	0,04	0,34	0,02	0,15	0,33	0,42
G4	0,17	0,03	0,26	0,02	0,10	0,30	0,34

<i>P[lArB]-PI</i> siRNA/piRNA ratio B	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	0,82	0,03	1,16	0,13	0,85	1,52	1,07
G2	0,21	0,02	0,37	0,03	0,14	0,49	0,29
G3	0,24	0,03	0,36	0,03	0,19	0,52	0,37

<i>P[lArB]-MI</i> siRNA/piRNA ratio I	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
G1	0,05	0,01	0,06	0,01	0,04	0,07	0,10
G2	0,08	0,03	0,11	0,01	0,07	0,12	0,17
G3	0,09	0,03	0,11	0,01	0,08	0,17	0,19
G4	0,10	0,02	0,12	0,01	0,07	0,19	0,21

<i>P-1152w^c</i>	<i>P[lArB]</i>	<i>DomF</i>	<i>lacZ</i>	<i>Hsp70</i>	<i>Adh</i>	<i>rosy</i>	<i>pBluescript</i>
	0,05	0,02	0,06	0,01	0,04	0,06	0,09

Conclusion

Dans cet article, nous avons montré qu'un transgène *P-lacZ* (*P[lArB]*) apporté par voie paternelle acquière au bout de quatre générations les mêmes propriétés répressives qu'un transgène apporté maternellement. Ceci s'accompagne en parallèle d'une augmentation de la production de piRNA au cours des générations. L'analyse plus fine des piRNA issu du transgène a permis de montrer que le transcrit du transgène passe d'un état d'ARNm à celui de précurseur de piRNA par l'effacement de l'épissage des introns. En effet, les introns présents dans le transgène finissent produisent des piRNA au même titre que les exons.

Le locus TAS peut donc transmettre ses propriétés répressives à une séquence qui s'y insère. Cette répression peut s'établir en quelques générations et suggère la transmission d'une mémoire épigénétique d'une génération à l'autre. Ce phénomène pourrait être généralisé aux autres clusters producteurs de piRNA présents dans le génome. L'arrivée d'un nouvel élément transposable dans un cluster de piRNA déjà existant permettrait la mise en place rapide mais pas immédiate de sa répression par l'effacement de l'identité de cet élément par le complexe RDC.

5. Régulation de l'expression du génome par les piRNA du locus TAS

A. Hypothèse de travail

Le rôle le plus connu des piRNA est de réprimer l'expression des éléments transposables dans la lignée germinale chez la drosophile (Brennecke *et al.* 2007), (Gunawardane *et al.* 2007), (Malone *et al.* 2009). Cependant de plus en plus d'articles aujourd'hui leur accordent aussi un rôle dans la régulation de gènes et ceux dans divers organismes comme nous avons pu le voir dans l'introduction (Rouget *et al.* 2010), (Kiuchi *et al.* 2014), (Gou *et al.* 2014).

Au cours de cette thèse, un des projets concernant la caractérisation fonctionnelle des TAS a été leur étude dans l'implication de la régulation de gènes. En effet, les locus TAS font partis des locus producteurs de piRNA. Il a été montré que lorsqu'un élément *P* est inséré dans le *TAS-X* celui-ci peut réprimer l'expression d'autres copies euchromatiques de *P*. Cette répression passe par les piRNA et dépend d'un mécanisme de type TSE impliquant une identité de séquences entre la source des piRNA et la cible. Sachant que les TAS sont des locus producteurs de piRNA, l'hypothèse (Figure 29) sur laquelle se base ce projet est donc la suivante, serait-il possible que des gènes possédant une séquence homologue aux TAS soient la cible de piRNA produit par les locus TAS en lignée germinale ?

Ce travail s'est articulé au départ autour de deux grandes questions. Premièrement, existe-t-il des gènes candidats possédant des fragments de TAS-R dans leur séquence. Et dans un deuxième temps, des gènes peuvent-ils être différentiellement exprimés en l'absence de séquence TAS-R dans le génome ?

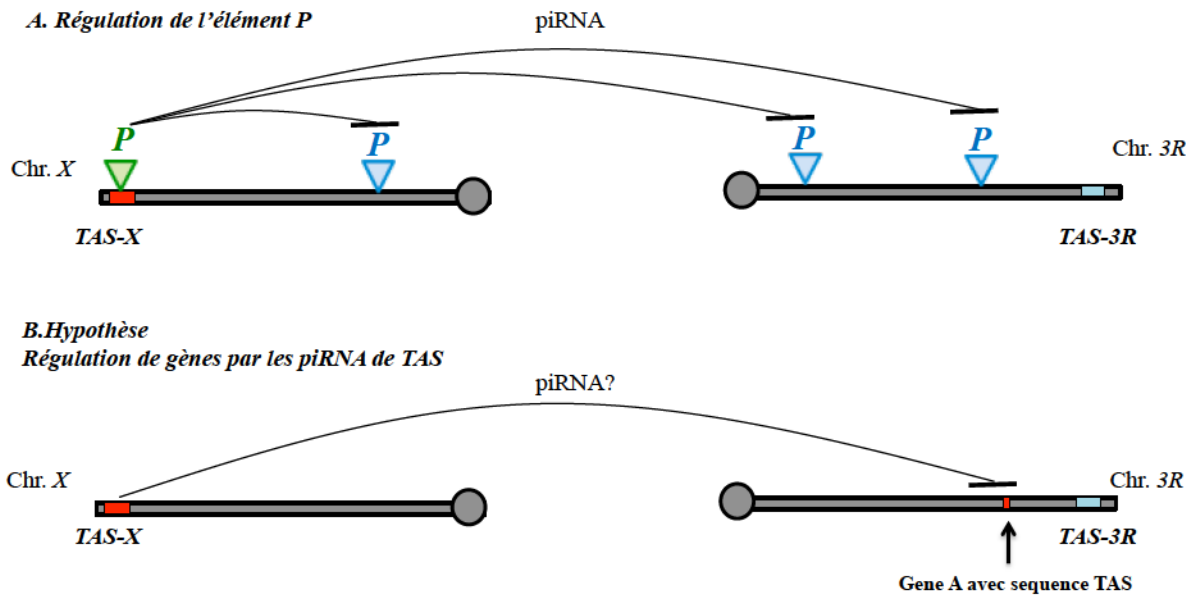


Figure 29. Régulation par les piRNA de l'élément P et de gènes. **A.** Régulation de l'élément P. L'élément P inséré dans le TAS-X peut produire des piRNA capables de réprimer l'expression d'éléments P euchromatiques. **B.** Régulation de gènes par les piRNA de TAS. Le locus TAS-X est capable de produire des piRNA. Certains gènes pourraient être la cible de ces piRNA s'ils possèdent une séquence homologue aux piRNA issus du locus TAS. Les cercles gris représentent les centromères. Le choix de représentation du chromosome 3 et du TAS-3R est arbitraire.

B. Recherche de gènes candidats potentiellement ciblés par les TAS-R

Pour la recherche de gènes candidats deux types des catégories ont été retenus. Les gènes qui possèdent un fragment de TAS dans leur transcrit appelés gènes avec TAS intra-géniques. Ces gènes pourraient être réprimés par homologie de séquence due aux piRNA de TAS. L'autre catégorie regroupe les gènes possédant une séquence TAS dans leurs régions flanquantes, dans une fenêtre de 3kb en amont et en aval, appelés gènes avec TAS flanquants. Dans ce dernier cas, la régulation du gène passerait par une hétérochromatinisation de ces régions flanquantes. Il a en effet été montré que la répression des cibles est accompagnée par l'établissement de marques répressives (H3K9me3) (Le Thomas *et al.* 2014) On peut supposer que cette marque pourrait se propager et réprimer l'expression d'un gène ayant dans ces régions flanquantes une séquence homologue aux TAS (Olovnikov *et al.* 2013).

B1. Analyse *in silico*

La première partie de ce travail s'est articulée autour de la recherche de fragments de TAS dans le génome de *D.melanogaster* ailleurs que dans les régions sub-télomériques. J'ai donc effectué une recherche de gènes candidats possédant des fragments de TAS dans leur séquence par des analyses bio-informatiques. Pour cela, l'outil BLAST (<http://blast.ncbi.nlm.nih.gov>) a été utilisé, et plus particulièrement le programme blastn plus performant pour l'alignement de séquences courtes (Figure 30). Celui-ci paraît adapté dans ce cas car nous recherchons des séquences pouvant correspondre à des piRNA de TAS dont la taille peut varier de 23-29nt. Dans les paramètres, la taille minimum de l'alignement considéré a été fixé à 11 nucléotides (word size), un mésappariement a été toléré et le filtre de faible complexité des séquences (low complexity filter) a été retiré car les TAS sont des régions riches en AT. Dans l'article montrant l'implication des piRNA du gène *nanos* (Rouget *et al.* 2010), la taille des séquences de piRNA était située entre 16 et 11nt pour la plus petite. Nous nous sommes donc basés sur ces données pour la recherche de gènes candidats. Afin de simplifier cette recherche, les séquences TAS ont été découpées en deux parties : les séquences correspondants au LTR *d'invader 4* et les séquences spécifiques à chacun des 3 TAS (*TAS-X*, *TAS-2R*, *TAS-3R*). Ce travail ayant été initié au laboratoire bien avant celui qui a permis d'identifier les différents domaines constituant les TAS (voir article I), nous avons utilisé toute la partie spécifique à chaque télomère des TAS.

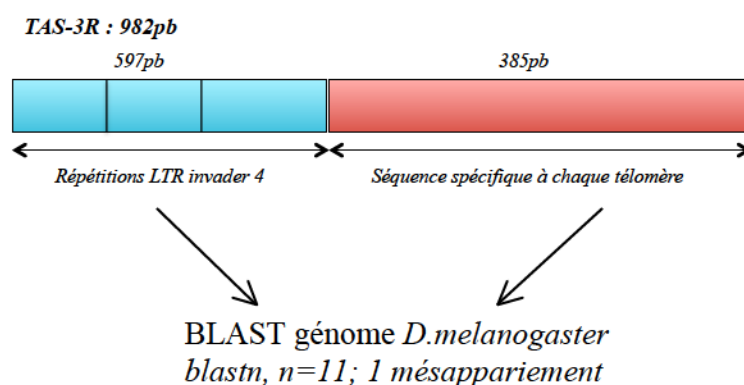


Figure 30. Paramètres utilisés pour la recherche de gènes candidats possédant une séquence TAS. Cette analyse a été effectuée avec les trois TAS du chromosome X, 2R et 3R. Le *TAS-3R* est utilisé à titre d'exemple.

B2. Les séquences TAS intra-géniques

Au total 351 gènes candidats possédant un fragment de TAS intra-génique (tous TAS confondus) ont pu être identifiés représentant au total 2,4% des gènes chez *Drosophila melanogaster*. Parmi ces gènes, 24 possèdent un fragment correspondant à *Invader 4* (majoritairement dans les introns). Ce sont des gènes pour lesquels on retrouve un fragment correspondant aux TAS au sein de leur séquence. Il peut s'agir d'exons codants ou non codants et d'introns. Plus de 80 % des gènes candidats possèdent un fragment de TAS dans les régions introniques et 20% dans les exons (Figure 31). Parmi les exons, on retrouve des fragments de TAS plus souvent dans les exons non codants, les 3' et 5' UTR (Figure 32). Par ailleurs, la taille des fragments de TAS que l'on retrouve majoritairement au sein des gènes est inférieure à 30pb (Figure 33). Cette taille pourrait correspondre à celle d'un piRNA (23-29nt). Les gènes possédant des fragments de TAS dont la taille est supérieure à 120pb sont des fragments correspondant à *Invader 4*.

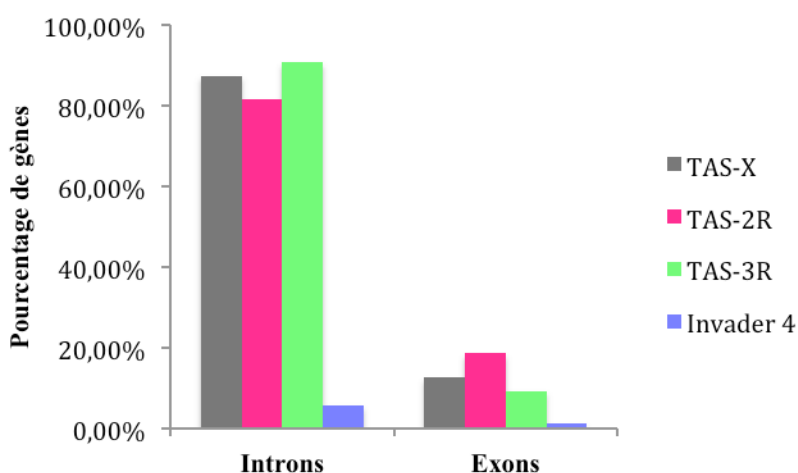


Figure 31. Pourcentage de gènes possédant une séquence *TAS-X*, *TAS-2R* ou *TAS-3R* dans les introns ou les exons. La catégorie exons comprend les exons codants et non codants (3'UTR et 5'UTR)

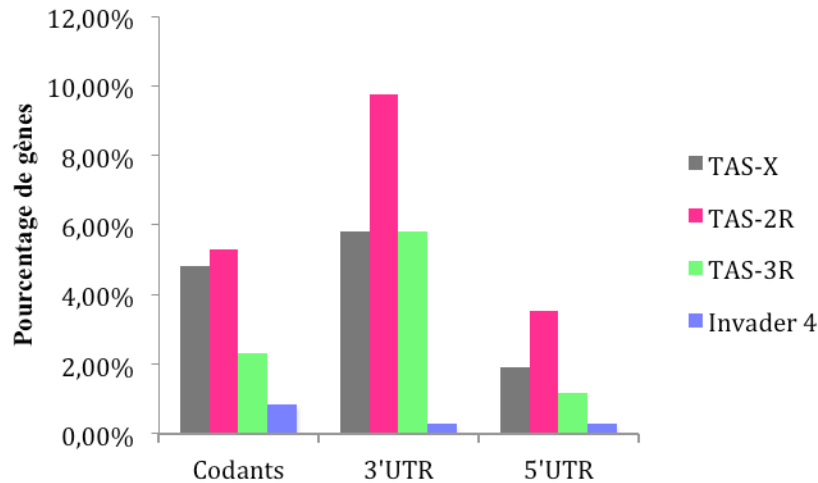


Figure 32. Répartitions des séquences *TAS-X*, *TAS-2R* et *TAS-3R* au sein de la catégorie exons.

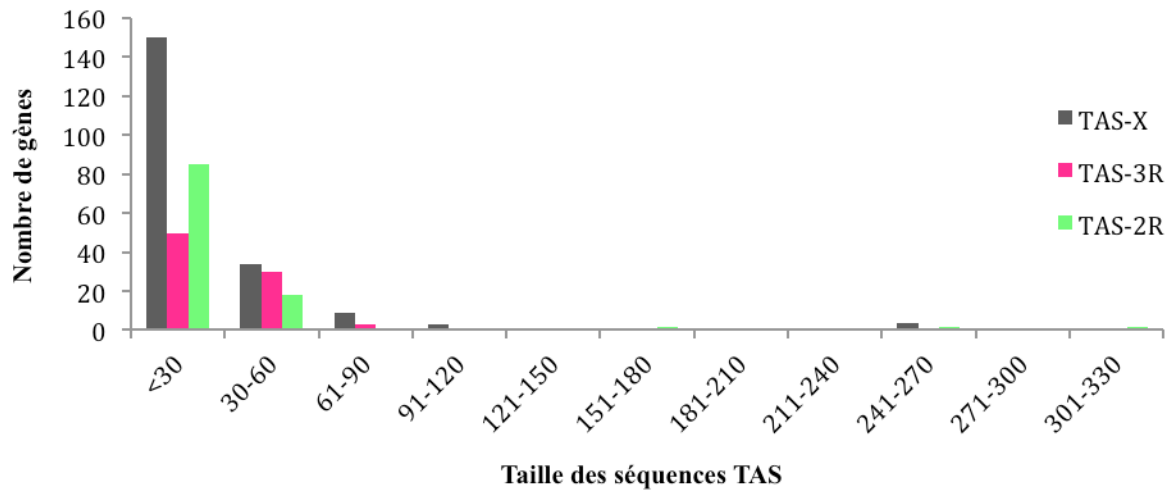


Figure 33. Tailles des séquences *TAS-X*, *TAS-2R* et *TAS-3R* retrouvées parmi les gènes.

B2. Les séquences TAS flanquantes aux gènes

Des fragments de TAS dans les régions flanquantes aux gènes ont aussi pu être trouvés. Ces fragments ont été recherchés dans une fenêtre de 3kb en amont et en aval des gènes. Au total 234 gènes (tous TAS confondus) entrent dans cette catégorie. Pour une grande majorité de ces fragments de TAS, leur taille est inférieure à 30pb et peut aller jusqu'à 60pb (Figure 34). D'autre part, aucun biais de présence de fragments de TAS en 5' ou en 3' des gènes n'a été retrouvé.

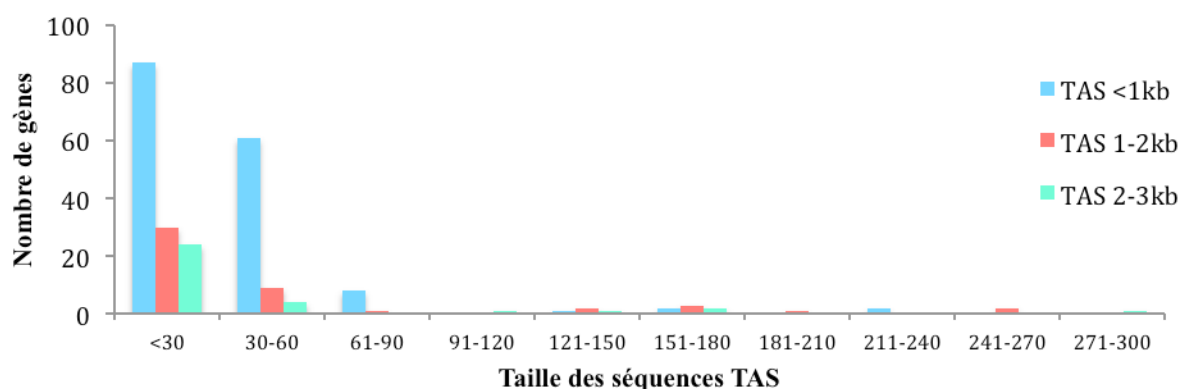


Figure 34. Tailles des séquences *TAS-X*, *TAS-2R* et *TAS-3R* retrouvées dans les régions flanquantes des gènes 3kb en amont et en aval.

Cette analyse a permis de trouver 585 gènes candidats possédant un fragment de TAS intra-génique ou flanquant. Certains de ces gènes ont ensuite été testés par RT-PCR quantitative (qRT-PCR).

C. Analyse par qRT-PCR

Pour tester l'expression par qRT-PCR de transcrits de gènes possédant un fragment de TAS intra-génique, un mutant pour le gène *Aubergine* un acteur de la voie des piRNA a été utilisé. L'hypothèse est que si un gène possédant un fragment de TAS est régulé par les piRNA du locus TAS, alors celui ci devrait être surexprimé en conditions mutantes *Aubergine*. En effet, la protéine Aubergine intervient lors de la boucle d'amplification « ping-pong » des piRNA (Brennecke *et al.* 2007) permettant la dégradation des transcrits d'ET. Si cette protéine est mutée, alors une accumulation des transcrits sera observée.

Le premier gène testé est *pep* (Protein on ecdysone puffs). Ce gène possède un fragment de TAS du domaine F spécifiquement présent dans une forme de ce transcrit dans un exon non codant. L'analyse par qRT-PCR montre que la forme possédant le fragment de TAS est surexprimée en condition mutantes alors que l'analyse d'un exon commun à toutes les formes de transcrits ne montre aucune différence quelque soit les conditions (Figure 35).

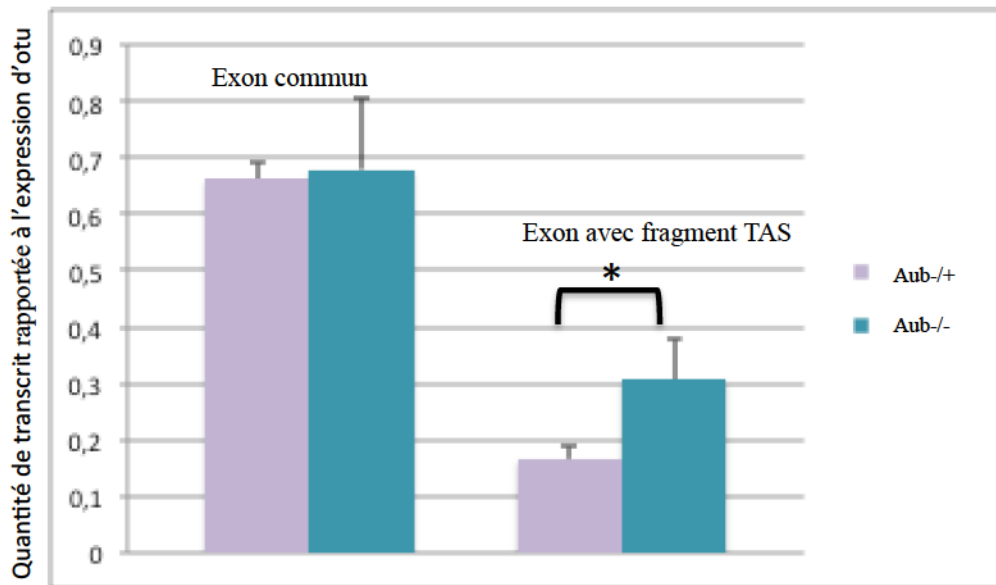


Figure 35. Expression du transcript de *pep* possédant un fragment de TAS dans un exon non codant comparé à un exon commun à toute les isoformes de ce transcript. La quantité est rapportée au gène *otu*.

Par ailleurs, *pep* possède un autre fragment de TAS présent dans un intron et correspondant au domaine E. Des amorces spécifiques de cette séquence ont été choisies et l'analyse par qRT-PCR montre que ces formes sont surexprimées en conditions mutantes pour Aubergine (Figure 36).

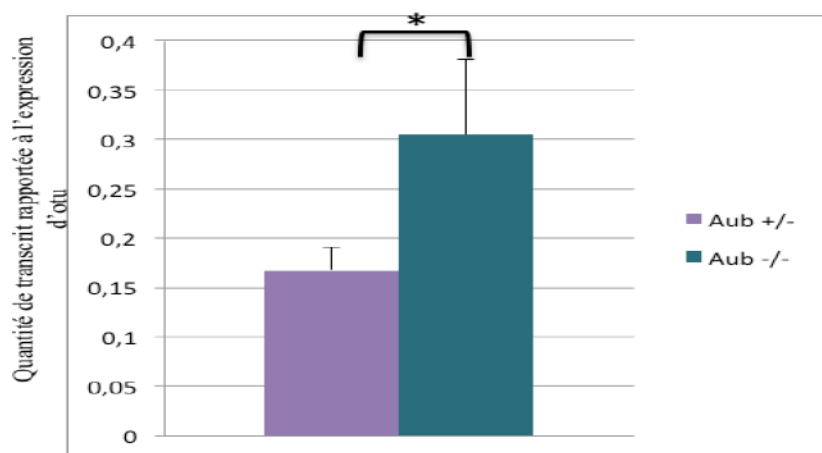


Figure 36. Expression du transcript de *pep* possédant un fragment de TAS dans un intron

Enfin je vous présente un dernier gène, *CG17219*, possédant un fragment de TAS correspondant au domaine G présent dans un exon de ce gène. L'expression de l'exon testé augmente en conditions mutantes pour Aubergine alors celle du gène présent dans le

voisinage de *CG17219* (*CG3605*) ne montre aucun changement d'expression quelque soit la condition (Figure 37).

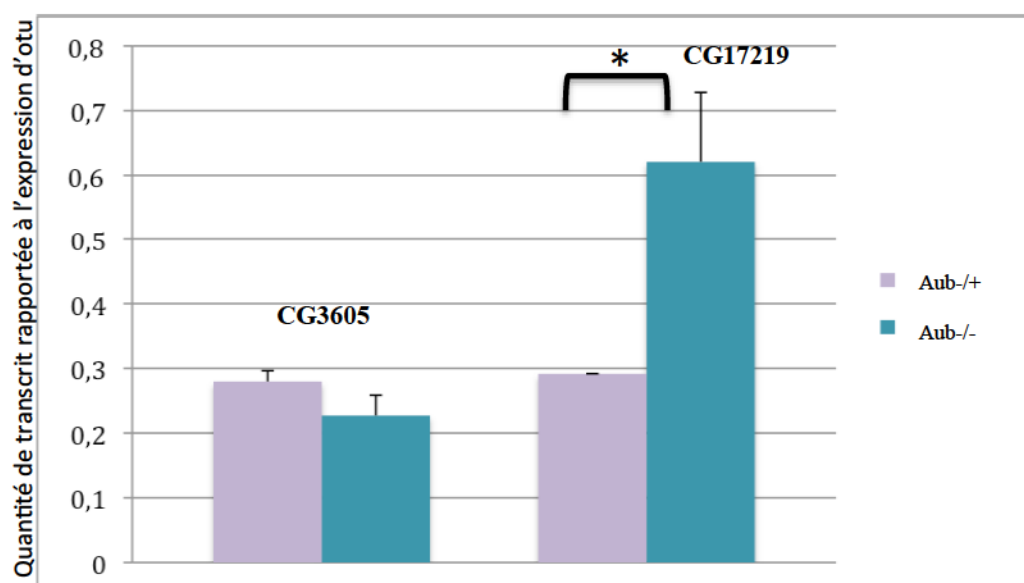


Figure 37. Expression du transcrit de *CG17219* possédant un fragment de TAS dans un exon et du gène *CG3605* présent en 5' de *CG17219*.

Ces premiers résultats étant assez encourageants, nous avons donc pu envisagé une analyse à plus grande échelle par l'étude du transcriptome par RNA-seq dans une souche sans TAS.

D. Construction de lignées sans TAS

J'ai pu caractériser par des expériences d'hybridation *in situ* sur chromosomes polytènes de drosophiles, plusieurs souches au laboratoire pour la présence des *TAS-X*, *2R* et *3R* (voir article I). En se basant sur les différentes combinaisons possibles pour la présence de TAS, nous avons tenté de construire une souche dépourvue de séquences TAS-R aux trois télomères grâce aux outils de la génétique de la drosophile. Cette souche représente la souche témoin idéale (appelée TAS-) afin de tester notre hypothèse de la régulation de gènes par les piRNA de TAS. Pour cela, une souche possédant le *TAS-X* et *TAS-2R* et non le *TAS-3R* ainsi qu'une souche possédant le *TAS-X* et *TAS-3R* et non le *TAS-2R* ont été utilisés. Les chromosomes sans *TAS-X*, *TAS-2R* et *TAS-3R* ont essayé d'être récupérés à l'aide de chromosomes balanciers (chromosomes remaniés permettant d'éviter la recombinaison et portant un marqueur

phénotypique). Malgré plusieurs tentatives aucune souche TAS- n'a pu être construite. Une hypothèse possible est que les chromosomes balanciers utilisés ne couvrent pas la totalité du chromosome et notamment les télomères. Des événements de recombinaisons seraient possibles dans la région télomérique. Une hypothèse alternative est que la souche sans TAS n'est pas viable, pour la plupart des lignées construites, aucune drosophile à l'état homozygote n'a pu être récupérée lors des constructions.

D'après les données présentées dans l'article I, il existe parmi TAS-R des petits domaines en communs entre les *TAS-X*, *TAS-2R* et *TAS-3R*. Un seul domaine est unique parmi ces TAS, le domaine F. Sa présence peut être contrôlée de deux manières. Par PCR à l'aide de couples d'amorces spécifiques de ce domaine, et par des expériences d'hybridation *in situ* sur chromosomes polytènes où les souches possédant un *TAS-X* (et donc un domaine F) peuvent être identifiées par des expériences d'hybridation *in situ* sur chromosomes polytènes. Le sujet s'est donc orienté à ce moment là sur le rôle des piRNA du *TAS-X* et plus précisément ceux issus du domaine F sur la régulation de l'expression du génome.

E. Les piRNA du domaine F peuvent-ils éteindre une cible ?

E1. Construction des transgènes et descriptions des lignées utilisées

Afin de valider fonctionnellement le rôle des piRNA du domaine F sur la régulation de l'expression d'une cible, j'ai construit un transgène cible *pUASp-RFP-DomaineF-K10*. Dans ce transgène, le gène codant la RFP (*Red Fluorescent Protein*) est en fusion transcriptionnelle avec le domaine F, inséré dans les deux sens de clonage possible (Figure 38). Ce transgène est sous le contrôle d'une séquence UAS activable par la protéine GAL4 qui est exprimée grâce au promoteur du gène *nanos* dans les ovaires. Il s'agit ci, d'une séquence UASp permettant une meilleure expression du transgène en lignée germinale et possédant 582pb correspondant à l'élément *P* (promoteur et une partie de l'exon 0). Enfin, la séquence K10 est le terminateur de transcription. Au total sept lignées transgéniques ont été obtenues dont quatre avec l'insertion du domaine F dans le sens 1 et trois avec l'insertion du domaine F dans le sens 2.

Deux souches disponibles au laboratoire ont été utilisées comme source pour apporté la protéine GAL4 (32180) afin d'activer l'expression du transgène *pUASp-RFP-DomaineF-K10*. Parmi ces deux souches, une possède le domaine F (*P-1152w^c* ; 32180 *BQ16*) et l'autre non

(w^{1118} ; 32180 *BQ16*) D'autre part, un cible *P-lacZ* (*BQ16*) est aussi présente dans les deux cas. Ce sont ces deux souches qui vont être utilisées comme source de piRNA (ou non) du domaine F.

L'hypothèse est que si les piRNA issus du domaine F télomérique ciblent la séquence du domaine F contenu dans le transgène, par complémentarité de séquence, l'expression de la RFP devrait être éteinte grâce au TSE.

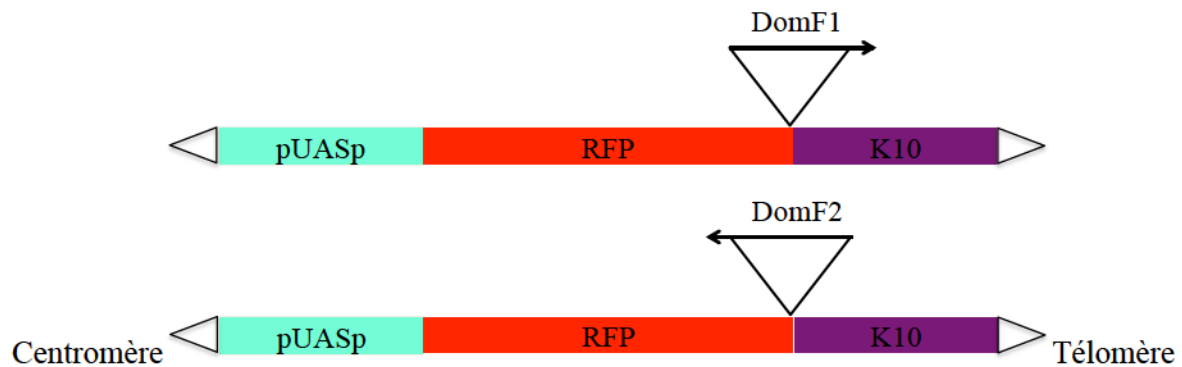


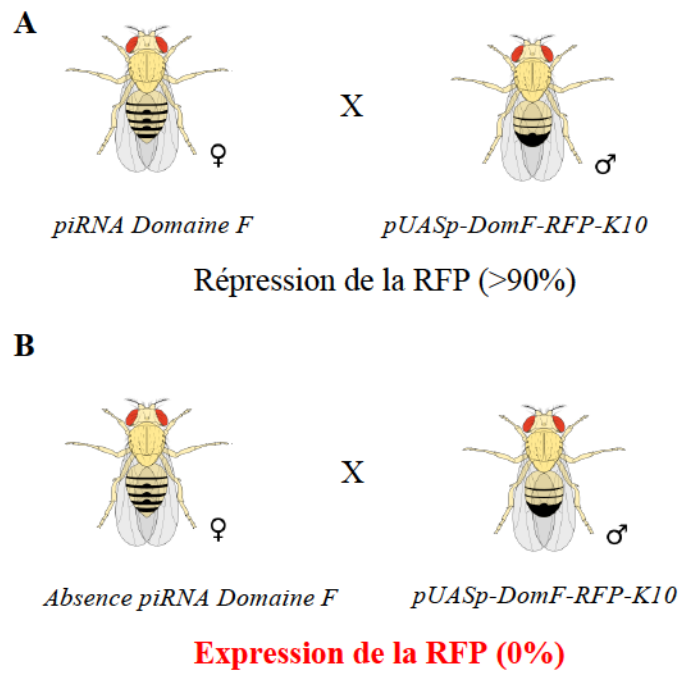
Figure 38. Structure des transgènes *pUASp-RFP-DomaineF*. Les deux types de transgènes obtenus après le clonage sont représentés sur le schéma. Le premier possède le domaine F cloné dans un sens (appelé DomF1). Le second possède le domaine F cloné dans le sens inverse (DomF2). Les sens 1 et 2 sont arbitrairement représentés allant du télomère au centromère

E2. Validation *in vivo* des lignées *pUASp-RFP-DomaineF*

E2.1 Principe de l'expérience

Les séries d'expériences qui vont suivre sont basées sur une des propriétés du TSE, l'effet maternel. En effet, lorsque la source de piRNA, homologue au transgène cible, est apportée par la femelle (Figure 39A), le pourcentage de répression est élevé (>90%). A l'inverse lorsque la source de piRNA est apportée par le mâle (Figure 39C), le pourcentage de répression est plus faible (0-40%). Ces résultats ont été observés par l'analyse du TSE lors d'utilisation de cibles *P-lacZ*. Par ailleurs, dans les cas où les piRNA du domaine F ne seront apportés ni par voie maternelle ni par voie paternelle, la RFP devrait être exprimée (Figure 39B et 39D). L'analyse de l'ensemble de ces résultats permettra de valider si la répression du transgène *pUASp-DomF-RFP-K10* dépend du TSE. Les résultats théoriques sont représentés Figure 39.

***pUASp-DomF-RFP-K10* apporté par le mâle**



***pUASp-DomF-RFP-K10* apporté par la**

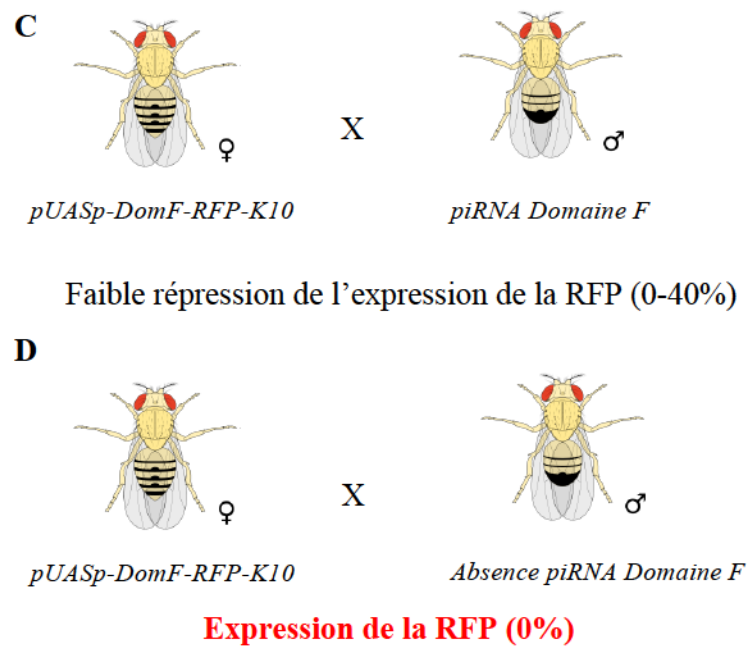


Figure 39. Résultats attendus dans le cas où le transgène *pUASp-DomF-RFP-K10* est transmis par le mâle ou par la femelle selon l'apport ou non de piRNA du domaine F. Les valeurs théoriques du TSE sont indiquées entre parenthèses.

E2.2. Expression du transgène *pUASp-RFP-K10* sans le domaine F

Avant d'analyser les transgènes obtenus dans les différents contextes (avec ou sans piRNA du domaine F), j'ai d'abord testé l'expression du transgène *pUASp-RFP-K10* sans l'insertion du domaine F par clonage en contrôle. Afin de vérifier si sans la présence du domaine F dans le transgène, des piRNA produits par les souches *w¹¹¹⁸; 32180 BQ16* et d'autre part la souche *P-1152w^c; 32180 BQ16* peuvent éteindre le *pUASp-RFP-K10*. Pour cela, la souche *pUASp-RFP-K10* a été croisé dans les deux sens d'une part avec la souche *w¹¹¹⁸; 32180 BQ16* et d'autre part la souche *P-1152w^c; 32180 BQ16*.

E2.2.1 Absence de piRNA du domaine F

Sans l'apport de piRNA du domaine F, l'expression de la RFP (en rouge), est observable dans les ovaires et ceci dans les deux sens de croisements (Figure 40A et Figure 40B). Le pourcentage de répression est de 0% (n=360) dans les deux cas. La souche *w¹¹¹⁸; 32180 BQ16* ne possède donc aucune séquence dans son génome produisant des piRNA homologues au transgène *pUASp-RFP-K10* induisant l'extinction de son expression.

E2.2.2 Avec piRNA du domaine F

La même expérience a été refaite cette fois-ci avec la souche *P-1152w^c; 32180 BQ16*. Celle-ci a été croisée dans les deux sens par la souche *pUASp-RFP-K10* (Figure 40C et 40D). Lorsque les piRNA du *TAS-X* sont apportés par la femelle, il y a une extinction de l'expression de la RFP (85%, n=280) (Figure 40C). Le transgène *pUASp-RFP-K10* possédant pas de séquence homologue avec le domaine F. Cette répression passe donc par l'homologie avec les séquences de l'élément *P* présentent à la fois dans le transgène *P-lacZ* télomérique et le transgène cible *pUASp-RFP-K10*. Les femelles souche *P-1152w^c; 32180 BQ16* synthétisent des piRNA déposés dans l'embryon permettant la répression du transgène. D'autre part, quand les piRNA de *P* sont apportés par le mâle, de la variegation peut être observée (Figure 40D). En effet, des chambres à œuf réprimées (blanches) et non réprimées (rouges) peuvent être distinguées. Le taux de répression dans ce cas est de 56,30% (n=360).

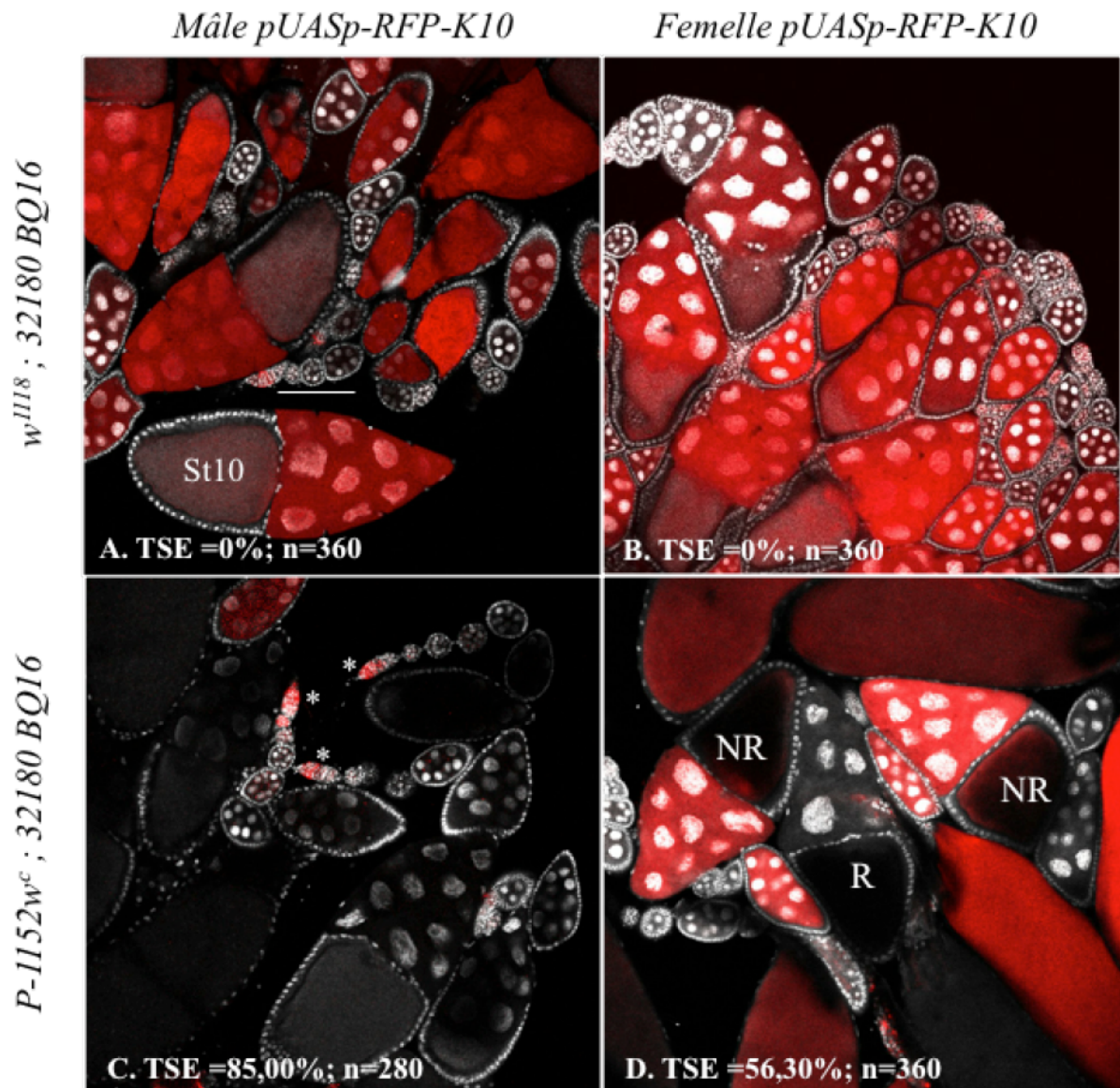


Figure 40. Expression du transgène *pUASp-RFP-K10* sans le domaine F croisé par les souches *P-1152w^c ; 32180 BQ16* et *w¹¹¹⁸ ; 32180 BQ16*. A. Expression de la RFP quand un cytoplasme vide de piRNA du domaine F est apporté par la femelle. On observe de la RFP en rouge dans la majorité des stades en lignée germinale. La RFP est d'ailleurs bien observable dans les cellules nourricières des stades 10 et non dans les cellules somatiques qui restent blanches. Par contre, les stades précoces semblent moins exprimer la RFP. Ceci est la conséquence d'une absence d'expression du GAL4 (32180) utilisé ici spécifiquement dans les stades 2-5 (symbolisés par un trait blanc). **B.** Expression de la RFP dans le sens de croisement réciproque. La RFP est observable comme dans le cas (A) et on retrouve la même absence d'expression dans les stades précoces. **C.** Extinction de l'expression de la RFP avec les piRNA issus de l'élément *P* présents dans les deux transgènes *P-lacZ* dans la souche *P-1152w^c*. Quelques germaniums non réprimés peuvent être distingués (symbolisés par des étoiles). **D.** Observation de la variegation de l'expression de la RFP. Trois stades 10 sont présentés, parmi eux un stade est réprimé (blanc) les deux autres sont rouges. Les pourcentages de répressions (TSE) ainsi que le nombre de chambre à œufs comptées (n) sont indiquées. Les observations ont été effectuées par microscopie confocale. Les ovaires ont été colorés au DAPI qui est représenté en blanc ici. NR : non réprimé ; R : réprimé

E2.3. Expression du transgène *pUASp-DomF-RFP-K10*

E2.3.1 Absence de piRNA du domaine F

Quand les piRNA du domaine F, ne sont pas apporté que ce soit paternellement (Figure 41A) ou maternellement (Figure 41B), dans les deux cas, il n'y a aucune répression de la RFP, les ovaires sont rouges. Les pourcentages de répression sont de 0% (n=240 et n=360 pour les Figure 41A et Figure 41B. respectivement). Cette expérience permet de souligner ici un élément très important. Il n'y a pas de piRNA du domaine F complémentaire à la séquence du domaine F dans le *pUASp-DomF-RFP-K10* capable d'éteindre son expression produite par un autre locus du génome. Cette observation exclurait aussi l'hypothèse que les piRNA du TAS-L, possédant une séquence commune avec le Dom F (la séquence *TLL*, voir article I), puissent intervenir dans cette répression. Certains piRNA communs entre ces deux séquences peuvent être attendues. En effet, les séquences *TLL* du domaine F et des TAS-L sont identiques à 80% (voir article I). Par l'analyse des banques de piRNA d'une souche *P-1152w^c* ainsi qu'une souche *w¹¹¹⁸* une explication est possible. La figure 42 montre que la souche *P-1152w^c* produit des piRNA du domaine F alors que *w¹¹¹⁸* n'est produit pas. Or *w¹¹¹⁸* possède des TAS-L et produit des piRNA de cette séquence (Figure 43). Cette analyse a été effectué en autorisant aucun mésappariement entre les séquences des piRNA du domaine F et les séquences de référence (Domain F et TAS-L). Si la même analyse en incluant cette fois-ci trois mésappariements possibles entre les piRNA du domaine F et la séquence du domaine F dans la souche *w¹¹¹⁸*. Dans ce cas, deux pics discrets correspondants à la séquence *TLL* du domaine F apparaissent alors que dans cette souche il n'y a pas de domaine F (Figure 44). Ces pics sont donc issus du TAS-L et sont alignés à cet endroit car les 3 mésappariements introduisent de la flexibilité lors de l'étape d'alignement. Cette analyse permet donc d'observer que deux séquences à priori très proches (80% identité), ne produisent pas forcément les mêmes piRNA. Ceci semble indiqué qu'une identité de séquence de 100% soit nécessaire entre le locus producteur de piRNA et le cible afin d'établir la répression. Les piRNA issus du DomF sont donc uniques car seul le *TAS-X* les produit.

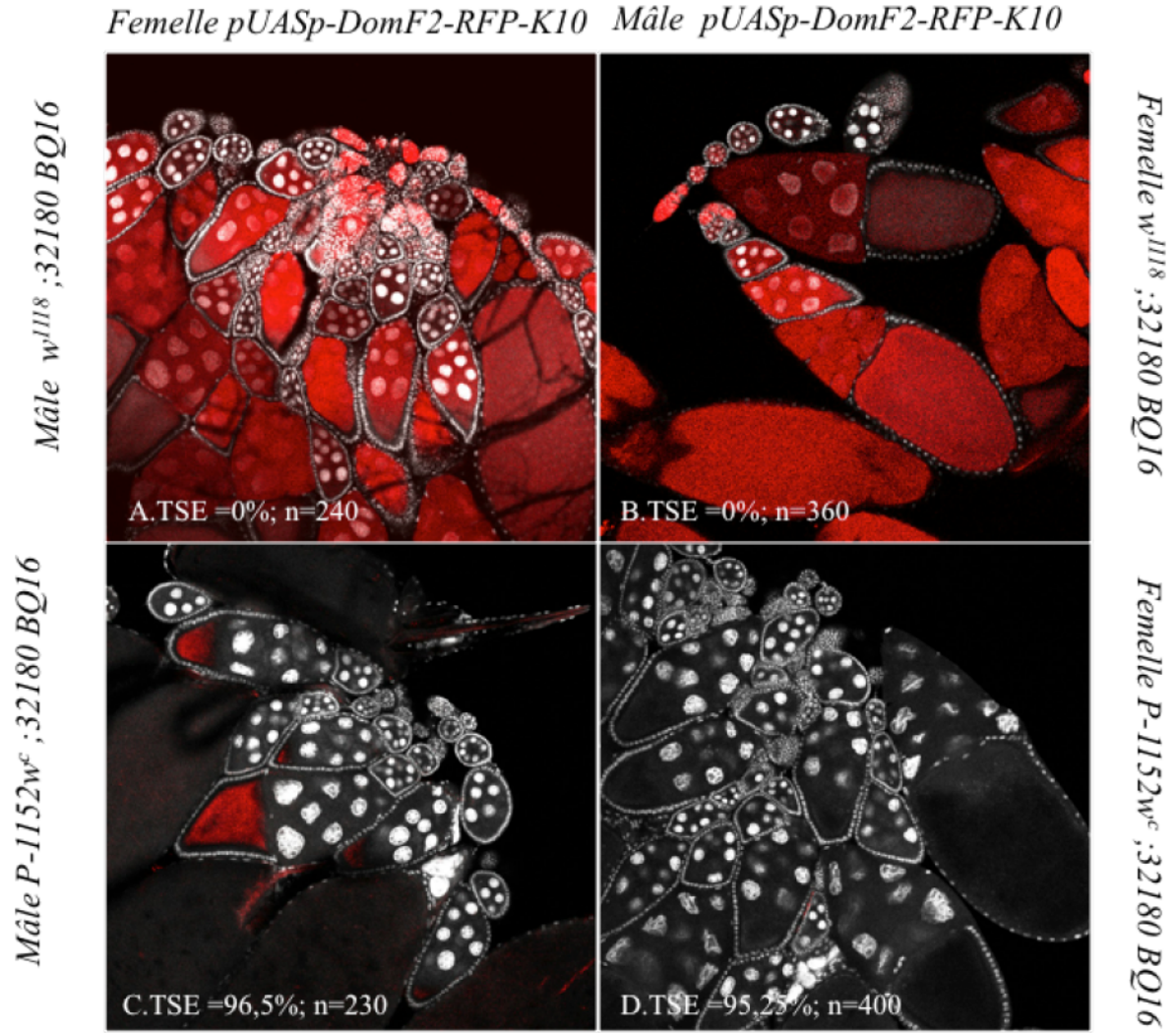


Figure 41. Expression du transgène *pUASp-DomF2-RFP-K10* croisé par les souches *P-1152w^c; 32180 BQ16* et *w¹¹¹⁸; 32180 BQ16*. **A-B.** Expression de le RFP dans les chambres à œufs. Les pourcentages de répressions (TSE) ainsi que le nombre de chambre à œufs comptées (n) sont indiquées. **C-D** Répression de le RFP dans les chambres à œufs. Les pourcentages de répressions (TSE) ainsi que le nombre de chambre à œufs comptées (n) sont indiquées. Le transgène *pUASp-DomF-RFP-K10* est inséré dans le sens 2 dans le cas présenté.

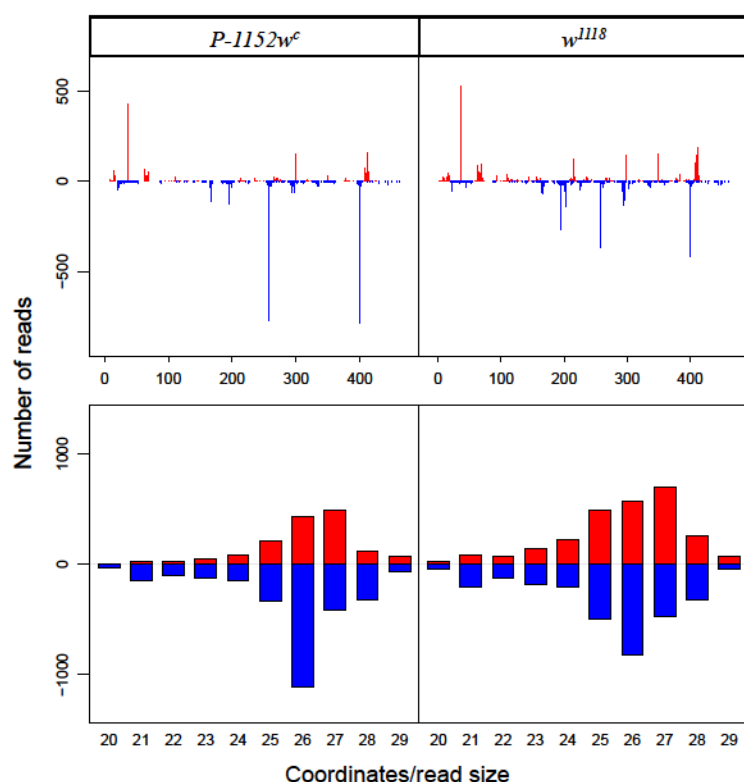


Figure 42. Répartition et taille des petits ARN du TAS-L dans une souche *P-1152w^c* ainsi qu'une souche *w¹¹¹⁸* avec 0 mésappariement possible lors de l'alignement des petits ARN sur le domaine F.

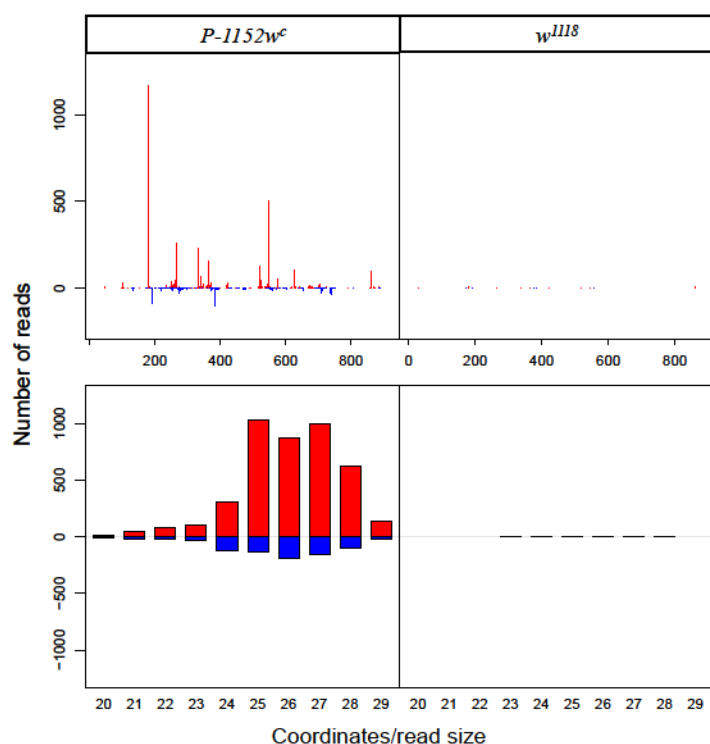


Figure 43. Répartition et taille des petits ARN du domaine F dans une souche *P-1152w^c* ainsi qu'une souche *w¹¹¹⁸* avec 0 mésappariement possible lors de l'alignement des petits ARN sur le domaine F.

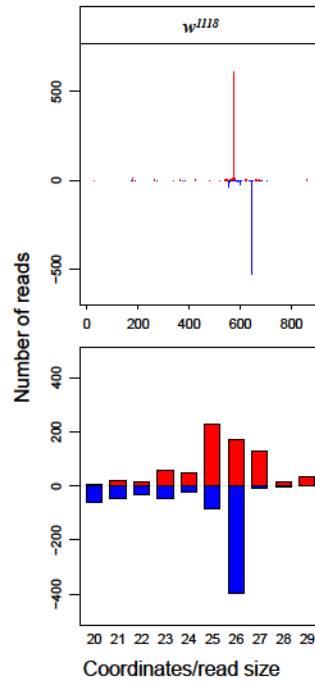


Figure 44. Répartition et taille des petits ARN du domaine F dans une souche w^{1118} avec 3 mésappariements possibles lors de l'alignement des petits ARN sur le domaine F.

E2.3.2 Avec piRNA du domaine F

Dans cette série d'expériences, les piRNA du domaine F ainsi que de l'élément *P*, produits par le transgène *P-lacZ*, sont transmis soit par le mâle (Figure 41C) soit par la femelle (Figure 41D).

En se basant sur les résultats obtenus précédemment (Figure 41A), une répression de l'expression de la RFP quand les piRNA sont transmis maternellement est attendue, à cause de la présence des piRNA de *P*. Dans le croisement inverse, c'est à dire quand les piRNA sont transmis paternellement, une répression d'environ 50% est attendue (Figure 41B). Ces résultats sont observés pour le premier cas (Figure 41D), le taux de répression du transgène est de 95,25% ($n=400$). Cependant, il y a plus de 96,50% ($n=230$) quand les piRNA du domaine F (et ceux de *P*) sont transmis par la mâle (Figure 41C). Ces résultats montrent que l'effet maternel n'est pas observé. Il semblerait que ce soit une combinatoire des piRNA issus de l'élément *P* ainsi que ceux issus du domaine F qui répriment le transgène *pUASp-DomF-RFP-K10*. Les mêmes résultats sont obtenus quelque soit le sens d'insertion du domaine F dans le transgène *pUASp-DomF-RFP-K10*.

E.2.4 Le TSE par une cible *P-lacZ*

Les souches utilisées ici, w^{1118} ; 32180 *BQ16* et $P-1152w^c$;32180 *BQ16*, possèdent un transgène *P-lacZ* cible, *BQ16*. Ce transgène ne possède pas de séquences communes avec le domaine F. Des sœurs des femelles issues de tout les croisements précédents ont été donc aussi testés pour la répression de cette cible par coloration au X-gal (Table 4).

	♀ w^{1118} ; 32180 <i>BQ16</i>	♂ w^{1118} ; 32180 <i>BQ16</i>	♀ $P-1152w^c$;32180 <i>BQ16</i>	♂ $P-1152w^c$;32180 <i>BQ16</i>
<i>pUASp-RFP-K10</i>	0% n=1000	0% n=1450	96% n=1300	31% n=1700
<i>pUASp-DomF2-RFP-K10</i>	0% n=800	0% n=500	97% n=1000	22,5% n=550
<i>pUASp-DomF1-RFP-K10</i>	0% n=850	0% n=600	99,9% n=1000	55% n=600

Table 4. Pourcentage de TSE obtenus pour les croisements présentés Figure 41 et Figure 45. Le nombre de chambres à œuf comptées est indiqué (n).

-La cible, *BQ16*, n'est pas réprimée (0%) quelque soit le sens de croisement et le transgène utilisé (*pUASp-RFP-K10* et *pUASp-DomF-RFP-K10*) quand aucun piRNA homologue à la cible est transmis (w^{1118} ; 32180 *BQ16*).

- Une forte répression de la cible est obtenue quand le transgène *P-lacZ* ($P-1152w^c$) est apporté par la femelle. Il y respectivement 96% et 97% de répression avec les transgènes *pUASp-RFP-K10* et *pUASp-DomF-RFP-K10*. Cette répression passe par les piRNA de l'élément *P* produit par la souche $P-1152w^c$. A l'inverse, quand le transgène *P-lacZ* est apporté par le mâle, de variegation ainsi qu'un faible taux de répression de la cible a été observé (31% et 22,5%). Il s'agit ici d'un cas typique d'effet maternel ce qui n'a pas été le cas dans les expériences précédentes concernant la répression du domaine F présente dans *pUASp-DomF-RFP-K10*.

Au sein des mêmes individus, deux types de répression ont donc pu être observées selon les séquences homologues au transgène cible utilisé. L'une présente un effet maternel (*P-lacZ* télomérique), l'autre n'en présente pas (les TAS).

Les piRNA du domaine F semble donc à priori être capable de réprimer l'expression d'une cible. Il faut cependant refaire ces expériences des souches où le driver Gal4 est directement présent dans cette souche *pUASp-DomF-RFP-K10*; 32180 afin de pouvoir tester des souches totalement vide d'élément *P* et possédant un *TAS-X*. Ceci permettra d'analyser seulement les piRNA issus du domaine F.

Il faut aussi noter que ces expériences (coloration au X-gal et expression de le RFP) ont été effectuées sur les 7 lignées possédant le transgène *pUASp-DomF-RFP-K10*. Il a donc pu être testé dans les deux sens d'insertions obtenus après le clonage (Figure 38). Les mêmes résultats ont été obtenus quelque soit le sens d'insertion du transgène (Figure 41 et Figure 45). Cela semble indiquer que le sens d'insertion du domaine F au sein du transgène n'a pas d'effet sur la répression. Les piRNA à la fois sens et anti-sens produits à partir du domaine F peuvent donc établir cette répression.

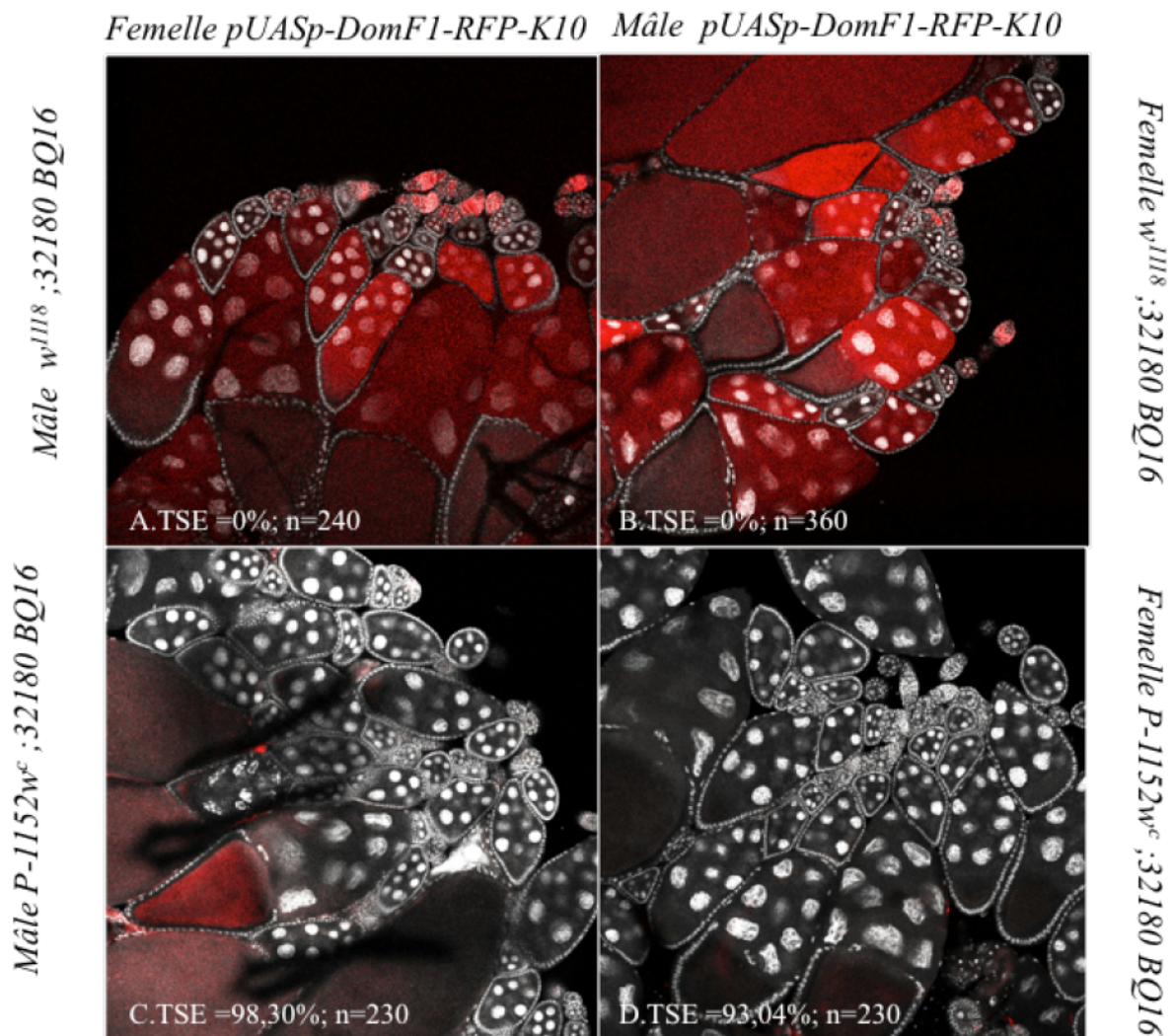


Figure 45. Expression du transgène *pUASp-DomF1-RFP-K10* croisé par les souches *P-1152w^c; 32180 BQ16* et *w¹¹¹⁸; 32180 BQ16*. A-B. Expression de le RFP dans les chambres à œufs. Les pourcentages de répressions (TSE) ainsi que le nombre de chambre à œufs comptées (n) sont indiquées. C-D Répression de le RFP dans les chambres à œufs. Les pourcentages de répressions (TSE) ainsi que le nombre de chambre à œufs comptées (n) sont indiquées. Le transgène *pUASp-DomF-RFP-K10* est inséré dans le sens 1.

Les résultats obtenus avec le transgène *pUASp-DomF-RFP-K10 ; 32180* montrent que les piRNA du domaine F sont fonctionnels. Une analyse du transcriptome a donc été lancée basée la présence ou l'absence du domaine F (et donc sans piRNA de ce domaine).

F3. Rôle des piRNA du *TAS-X* dans l'expression du génome.

Une analyse par RNA-sequencing (RNA-seq) a été initiée au laboratoire afin de comprendre le rôle des piRNA du *TAS-X* dans l'expression du génome. Pour cela, nous nous sommes basés sur une propriété intrinsèque de la régulation par les piRNA (effet maternel) et sur l'existence de lignées possédant un *TAS-X* ou non. Ce sont les piRNA du domaine F qui seront plus particulièrement analysés car ce domaine est uniquement présent dans le *TAS-X*. Les souches ont été appelées DomF+ (présence de *TAS-X*) et DomF- (pas de *TAS-X*).

Des individus DomF+ et DomF- ont donc été croisés dans les 2 sens. Les ovaires des descendants ont été disséqués pour extraire les ARNm et les petits ARN. Ainsi les deux types de femelles séquencées ont le même génotype mais un épigénome différent dû à l'héritage maternel différentiel des piRNA du DomF permettant leur comparaison.

Suite à l'analyse des piRNA issus des femelles sœurs envoyées pour le RNA-seq, un résultat inattendu a été obtenu. En effet, l'équipe avait montré que la capacité répressive d'un transgène inséré dans les TAS s'acquiert progressivement au cours des générations (Josse et al, 2007). Nous pensions que le locus TAS possédait la même dynamique d'acquisition des propriétés de répression. Nous nous attendions ainsi à ce qu'il y ait plus de piRNA du DomF à la 1^{ère} génération pour le croisement ♀ DomF+ x ♂ DomF- que pour le croisement réciproque, ceci permettant de comparer les deux épigénomes. Or dès la 1^{ère} génération, les TAS produisent autant de piRNA du DomF dans les deux sens de croisement ce qui rend impossible la comparaison initialement prévue.

Toutefois, ces résultats nous ont ouverts une piste intéressante pour permettre de comprendre la dynamique d'acquisition de la production de piRNA par un locus lorsqu'un élément s'y insère. Une analyse plus poussée a été initiée en se basant sur les résultats obtenus et a permis d'avoir des données présentées dans l'article II.

En ce qui concerne, l'analyse du rôle des piRNA du *TAS-X* dans l'expression du génome, une approche plus classique du RNA-seq a été adoptée. Les lignées DomF⁺ et DomF⁻ ont été isogénisées par croisements successifs sur 5 générations et le RNA-seq a été refait.

F.3.1 Analyse du transcriptome des lignées isogénisées

F.3.1.1 Présentation des lignées

Lors de ce deuxième RNA-seq, les lignées utilisées DomF⁺ et DomF⁻ ont été isogénisées. La lignée appelée DomF⁺ est la souche *P-1152w^c* qui possède deux transgènes *P-lacZ* (*[P-lArB]*) insérés dans le *TAS-X*. Cette lignée a été utilisée car elle permet de s'assurer de la présence du *TAS-X* et donc du domaine F en suivant la présence des transgènes grâce au TSE. La lignée DomF⁻ est la souche *Canton^y* ne possédant pas de *TAS-X*. Suite à l'isogénisation, les capacités répressives des deux lignées ont été testées grâce au TSE afin de vérifier que cette étape n'a pas altéré la présence du transgène dans le *TAS-X*. Pour la souche *Canton^y* le pourcentage de TSE mesuré est de 0% n= 1100 et pour la souche *P-1152w^c* de 98,1% n=1400. D'autre part comme lors de la première campagne de RNA-seq, les ARN totaux extraits à partir d'ovaires de femelles des deux génotypes ont été envoyés d'une part au séquençage pour les petits ARN et d'autre part au séquençage du transcriptome. Les expériences ont été effectuées en triplicat. Nous avons analysé les petits ARN produits par le domaine F dans les deux cas. Pour les réplicats concernant la souche *P-1152w^c* qui des piRNA issus du domaine F peuvent être observés alors que pour la souche *Canton^y* en sont bien vides (Figure 46). Par ailleurs, l'analyse des piRNA du domaine F pour les triplicats *P-1152w^c* permet de montrer les mêmes profils de piRNA dans les trois cas. Nos contrôles sont bien validés. Le RNA-seq a donc été effectué dans les bonnes conditions avec une souche ne produisant pas de piRNA du domaine F (*Canton^y*) et une souche en produisant (*P-1152w^c*).

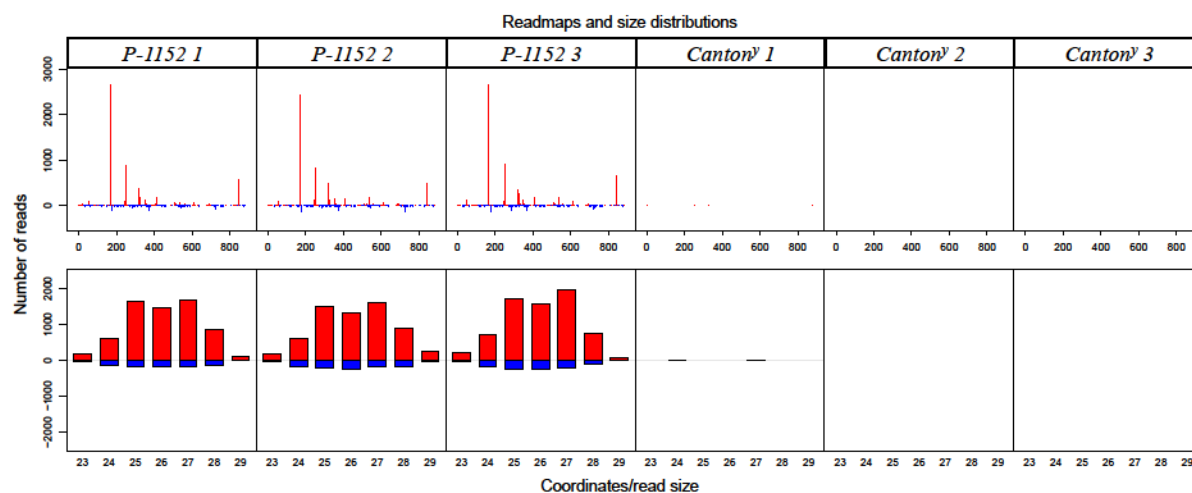


Figure 46. Distribution des piRNA issus des réplicats pour les souches utilisées pour les RNA-seq (*P-1152^{w^c}* et *Canton^y*) sur la séquence du domaine F.

F.3.1.2 Analyse préliminaire

Je vous présenterai dans cette partie une première analyse que j'ai effectuée suite à l'obtention des données de séquençage.

L'analyse des transcrits différentiellement exprimés entre la condition DomF- comparée à la condition DomF+ (DomF-/DomF+) nous a été délivrée par la plate-forme de séquence de l'Ecole Normale Supérieure où le RNA-seq a été effectué. C'est le programme STAR qui a été utilisé pour l'étape d'alignement sur le génome. Celui-ci à l'avantage d'aligner des séquences à cheval sur deux exons alors qu'avec d'autres programmes comme Bowtie par exemple ces séquences ne sont pas prises en compte.

Au total 1072 gènes sont différentiellement exprimés entre la condition DomF- comparée à la condition DomF+ avec une p-value ajustée inférieure à 0.05 ont été trouvés. Le transcrit qui est statistiquement le plus différentiellement exprimé entre ces deux conditions est celui d'*Adh*. Ce gène dont le transcrit est fortement exprimé dans les ovaires (données ModEncode), possède une p-value ajustée de 0 et est 12 fois plus surexprimés quand le domaine F n'est pas présent. En effet, la région correspondant au gène *Adh* est présente dans deux les transgènes *P-lacZ* insérés dans les *TAS-X* de la souche *P-1152^{w^c}* utilisée pour ce RNA-seq. Cette souche possède donc quatre copies du gène *Adh* en plus de la copie endogène. D'autre part, comme nous avons pu le voir dans l'article 2, la séquence *Adh* du transgène produit des piRNA dirigés contre ce gène. Ainsi dans la souche *P-1152^{w^c}*, le transcrit d'*Adh* est dégradé par la machinerie de la voie de biogenèse des piRNA et apparaît comme différentiellement exprimé dans cette analyse. De la même manière, *CG3281* qui est aussi présent dans le transgène *P-lacZ* télomérique fait parti des transcrits les plus

différentiellement exprimé d'un point de vue statistique avec une p-value ajustée de 1,24E-25 et est 2 fois plus surexprimé quand le domaine F n'est pas présent.

Parmi ces gènes, 634 sont surexprimés et 438 transcrits sont sous exprimés dans la souche DomF⁻ par rapport à la souche DomF⁺.

Les gènes qui nous intéressent dans un premier temps sont ceux qui sont surexprimés. En effet, selon notre hypothèse de régulation de gènes par les piRNA du domaine F, si un transcrit est la cible des piRNA du domaine F alors dans une souche DomF⁺ il sera plus faiblement exprimé, son expression étant réprimée par les piRNA du domaine F.

Dans un premier temps, j'ai donc recherché par BLAST, parmi les 634 gènes surexprimés ceux possédant des séquences pouvant s'aligner contre les piRNA du domaine F. Ces derniers ont récupérés en regroupant les trois banques séquençages des petits ARN dans les souches *P-1152w^c* (Figure 12).

Au total, 132 gènes possédant une séquence sur le même brin que le transcrit, que 126 gènes possédant une séquence sur le brin complémentaire du transcrit, ainsi que 126 gènes possédant une séquence de piRNA du domaine F à la fois sur les deux brins ont pu être trouvés.. La taille de ces séquences varie de 11 à 20 nucléotides sans aucun mésappariements. Pour la suite de l'analyse, c'est la deuxième catégorie de gènes qui nous intéresse. En effet, pour que les piRNA du domaine F puissent ciblés la séquence d'un transcrit ces derniers doivent justement être en anti-sens par rapport à ce transcrit. Dans un deuxième temps, les termes d'ontologie enrichis aussi appelés « Gene Ontology » (GO) (Ashburner et al, 2000) dans ces trois catégories de gène trouvés ont été cherchés (Table 1, Table 2 et Table 3). Ce type d'analyse est notamment utilisé lors de l'étude d'un nombre important de gènes issus par exemple d'analyses transcriptomiques. Celle-ci permettrait de mettre en évidence si certains processus biologiques sont enrichis dans les conditions mutantes par rapport aux conditions sauvages, reflétant ainsi la dérégulation d'une voie commune à ces gènes. Il existe plusieurs programmes permettant d'effectuer ce type d'analyse. Dans le cas présent, c'est l'outil DAVID (Huang et al, 2009) qui a été utilisé pour établir les termes de GO.

<i>Termes d'ontologie</i>	<i>GO:id</i>	<i>Nombre de gènes</i>	<i>P-Value</i>
multicellular organism reproduction	GO:0032504	19	3,60E-04
reproductive process in a multicellular organism	GO:0048609	19	3,60E-04
gamete generation	GO:0007276	18	3,90E-04
sexual reproduction	GO:0019953	18	5,40E-04
reproductive cellular process	GO:0048610	14	9,40E-04
regulation of cell morphogenesis	GO:0022604	7	1,00E-03
regulation of cell shape	GO:0008360	6	2,50E-03
vitelline membrane formation in chorion-containing eggshell	GO:0007305	3	3,20E-03
vitelline membrane formation	GO:0030704	3	3,20E-03
ovarian follicle cell development	GO:0030707	8	3,50E-03
regulation of development, heterochronic	GO:0040034	3	4,80E-03
germ cell migration	GO:0008354	4	5,80E-03
cell migration	GO:0016477	7	6,90E-03

Table 5. Terme d'ontologie enrichis pour les gènes sur exprimés possédant une séquence complémentaire de piRNA de domaine F sur le même brin que le transcrit du gène.

<i>Termes d'ontologie</i>	<i>GO:id</i>	<i>Nombre de gènes</i>	<i>P-Value</i>
ion transport	GO:0006811	9	1,80E-02
extracellular structure organization	GO:0043062	4	2,90E-02
axonogenesis	GO:0007409	6	4,30E-02
neuron differentiation	GO:0030182	9	4,40E-02
sensory organ development	GO:0007423	9	4,40E-02
neuron development	GO:0048666	8	5,10E-02

Table 6. Terme d'ontologie enrichis pour les gènes sur exprimés possédant une séquence complémentaire de piRNA de domaine F sur le brin complémentaire du transcrit du gène.

<i>Termes d'ontologie</i>	<i>GO:id</i>	<i>Nombre de gènes</i>	<i>P-Value</i>
cell-cell signaling	GO:0007267	8	1,70E-03
intracellular signaling cascade	GO:0035556	9	3,00E-03
protein amino acid phosphorylation	GO:0006468	8	5,50E-03
compound eye photoreceptor cell differentiation	GO:0001751	5	1,30E-02
eye photoreceptor cell differentiation	GO:0001754	5	1,60E-02
compound eye morphogenesis	GO:0001745	7	1,70E-02
photoreceptor cell differentiation	GO:0046530	5	2,20E-02
eye morphogenesis	GO:0048592	7	2,40E-02
regulation of cell size	GO:0008361	4	2,40E-02
multicellular organism reproduction	GO:0032504	13	2,50E-02

Table 7. Terme d'ontologie enrichis pour les gènes sur exprimés possédant une séquence complémentaire de piRNA de domaine F sur les deux brins transcrits.

L'analyse des termes GO permet d'observer un enrichissement pour les gènes impliqués dans les processus développementaux ainsi que la reproduction pour les gènes possédant une séquence de piRNA complémentaire au sens du transcrit (Table 5). Ces termes peuvent refléter les gènes exprimés dans le tissu ovarien qui a été utilisé pour l'étude du transcriptome.

Alors que pour les gènes possédant une séquence de piRNA issus du domaine F correspondant au brin complémentaire par rapport au gène montre plutôt un enrichissement pour les gènes impliqués dans les processus neuronaux (Table 6). Ces gènes pourraient être réprimés par des piRNA du domaine F au cours du développement. Alors que les gènes présentant un piRNA du domaine F pouvant s'apparier sur les deux brins du transcrit, ce sont les termes correspondant à la vision sont enrichis. Les listes des gènes utilisés pour les analyses de GO sont disponibles dans les annexes.

Il faudrait maintenant affiner cette analyse en regardant la localisation des séquences de piRNA du domaine F dans les transcrits plus spécifiquement. Par exemple, trouve t-on un enrichissement de ces séquences dans les exons ou les introns? D'autre part, une analyse du niveau d'expression de ces gènes dans la condition DomF- par rapport à la condition DomF+ par RT-PCR quantitative devra aussi être effectuée afin de valider les gènes trouvés.

Par ailleurs, il serait intéressant d'étudier plus en détails ces gènes notamment ceux impliqués dans les processus neuronaux. En effet, il a été montré que la domestication des drosophiles au laboratoire est accompagnée d'un changement de l'expression des gènes neuronaux (Stanley and Kulathinal 2016). Par ailleurs, il a aussi été montré qu'il existe des piRNA somatiques dans une population particulière de cellules constituant les « mushroom bodies » dans le cerveau de la drosophile (Perrat *et al.* 2013). Ces piRNA pourraient peut être participés dans la répression de certains gènes dans les cellules somatiques (le cerveau par exemple).

III- Discussion et perspectives

Les éléments transposables sont une partie intégrante des génomes. Ils peuvent être bénéfiques ou délétères pour l'organisme si leur transposition n'est pas contrôlée. Ainsi divers mécanismes permettant leur répression ont pu être mis en place par les organismes hôtes. Chez *Drosophila melanogaster*, une classe spécifique de petits ARN, les piRNA assurent la répression des ET en lignée germinale. Au cours de cette thèse, je me suis plus particulièrement intéressée, à un locus producteur de piRNA, le locus TAS.

Ce travail s'est articulé autour de deux projets, d'une part l'étude de la structure et de la distribution des TAS au sein des génomes, et d'autre part l'analyse de leur rôle fonctionnel.

Structure et distribution des TAS

Le premier objectif de cette thèse a été de caractériser de manière plus fine la structure des TAS ainsi que le polymorphisme pour leur présence dans différentes souches.

La structure des TAS

J'ai pu montrer que la famille des TAS-R est constituée, en plus de la présence du LTR d'*Invader 4*, de petits domaines d'ADN réarrangés différemment selon le télomère. Les séquences de ces domaines sont conservées et pourraient constituer des motifs pour la fixation de protéines comme des facteurs de transcription par exemple. Il serait intéressant de rechercher de manière plus poussée si certains motifs peuvent être détectés. Un facteur de transcription, Woc (Without Children), a notamment été retrouvé dans un crible de protéines qui pourraient être présents sur les TAS (Antao et al, 2012). De plus, cette protéine possède un domaine de fixation pour les séquences riches en AT, ce qui est le cas des TAS. Une souche Woc-GFP est disponible au laboratoire. Je l'ai utilisé afin de pouvoir co-localiser la protéine Woc sur le locus TAS à l'aide d'un anticorps anti-GFP. Les essais n'ont pas été concluants, il semblerait que cette souche n'exprime pas ou très faiblement la GFP. Des nouvelles expériences d'hybridation *in situ* en utilisant une sonde spécifique pour les TAS ainsi qu'un anticorps dirigé contre la protéine Woc devront être refaites. Ceci permettrait de confirmer que Woc est bien présent sur les TAS pouvant ainsi intervenir dans leur transcription.

D'autre part un domaine de 158pb a pu être mis en évidence, le *TLL*. Cette séquence est présente dans le domaine spécifique des *TAS-X*, le domaine F des *TAS-R* et est le constituant majeur des *TAS-L*. Le *TLL* est un lien qui nous a permis de rapprocher les deux familles de *TAS*. L'étude de cette séquence dans les autres espèces de drosophiles a mis en évidence que le *TLL* est restreint au groupe melanogaster. Des séquences correspondants au *TLL* ont été trouvées chez *D. sechellia*, *D. simulans*, *D. erecta* et *D. yakuba* dans les régions télomériques. La découverte de ce domaine permet de proposer une histoire évolutive des *TAS*. Des séquences de type *TAS-L* formées par des *TLL* ont pu être présentes à un moment donné et une transposition d'*Invader 4* parmi les *TAS-L* aurait permis l'émergence de la famille des *TAS-R* chez *D. melanogaster*. Par ailleurs, lors de la recherche des séquences *TLL*, un grand nombre d'entre elles ont été retrouvés dans des régions de chromosomes non assemblés chez *D. sechellia*, *D. simulans*, *D. erecta* et *D. yakuba*. Leur position télomérique ne peut être déterminé dans ce cas. Des expériences d'hybridations *in situ* en utilisant comme sonde les régions possédant des *TLL* de chacune des espèces permettrait de pouvoir les localiser.

La distribution des *TAS* dans les populations de drosophiles

L'autre axe développé dans cette thèse concerne l'étude de la distribution des *TAS* parmi des populations de drosophiles récoltées récemment dans la nature comparée aux souches conservées depuis plusieurs décennies dans les laboratoires. J'ai pu montrer qu'il n'existe aucune souche de *D. melanogaster* dépourvues de *TAS-R* mais qu'un polymorphisme peut être observé pour leur présence aux trois télomères (*X*, *2R* et *3R*). Au moins deux locus *TAS* ont été retrouvés dans chaque cas. Ceci concerne aussi bien les souches de la nature que les souches de laboratoire. La grande différence reste la présence systématique du *TAS-X* dans les populations naturelles alors qu'un maintien de ces souches dans les laboratoires entraîne progressivement sa perte. Ces résultats sont importants car jusqu'à présent la perte du *TAS-X* dans les laboratoires avait été suggérée mais jamais clairement démontré. La question soulevée par ces résultats est pourquoi une telle différence entre ces deux conditions de vie de la drosophile ?

Plusieurs hypothèses peuvent être émises. La première est que dans la nature les souches de drosophiles sont toutes envahies par l'élément *P*. Sachant qu'un point chaud d'insertion de cet élément est le *TAS-X* (Ronsseray et al, 1997), ce *TAS* pourrait donc être maintenu à cause de cette pression exercée par *P*. Celle-ci se traduit par une production de piRNA de *P* qui réprime les *P* euchromatiques assurant donc une descendance fertile. Il faut aussi souligner que dans les laboratoires, le *TAS-X* est retrouvé majoritairement dans les souches où justement

celui-ci été sélectionné pour sa présence car un élément *P* ou un transgène *y* est inséré. La perte des *TAS-X* pourrait être causée par des cassures terminales survenant aléatoirement au fil des générations. Les individus étant conservés dans un tube comme une « population homogène », cet événement finirait par se répercuter sur toutes les drosophiles après un temps *t*. Le résultat final serait donc la perte du *TAS-X*. Dans la nature, les populations sont plutôt hétérogènes et ceci permettrait donc le maintien de ce TAS. Parmi les toutes souches qui ont été récoltées dans la nature, 7 ont perdus leur *TAS-X* au bout de 2 ans. Il serait intéressant d'établir une cinétique pour les autres souches afin d'estimer le temps moyen de maintien au laboratoire conduisant à la perte.

Une dernière hypothèse concernerait le rôle du *TAS-X* dans la régulation de gènes. Les piRNA issus de ce locus seraient nécessaires dans la nature pour réprimer l'expression de gènes dispensables dans les conditions de laboratoires. En effet, dans le tube de culture, les drosophiles n'ont pas besoin de rechercher un partenaire sexuel, de la nourriture ou bien de s'orienter dans un vaste espace. Des différences comportementales entre drosophiles de la nature et celles maintenues au laboratoire ont récemment été décrites (Stanley et Kulathinal, 2016). Les auteurs montrent que ces différences résulteraient de mutations s'accumulant dans des gènes impliqués dans les processus neuronaux.

Fonctions des TAS

Le deuxième objectif de cette thèse a concerné l'étude du rôle fonctionnel des TAS au sein du génome de *D. melanogaster*.

Les TAS, un cluster de piRNA

J'ai montré que les TAS sont des locus producteurs de piRNA. Ceci concerne non seulement les LTR d'*Invader 4* mais aussi les régions flanquantes constituées par les différents petits domaines. L'analyse de mutants a permis de voir que les piRNA issus des TAS-R dépendent des protéines intervenant dans leur voie de biogenèse : Rhino, Cutoff, Zucchini et Aubergine. Leurs effets sur les TAS-L apparaît de façon moins claire notamment pour les mutants Aubergine et Zucchini. Ceci pourrait refléter une biologie différente pour la production des piRNA des TAS-R et des TAS-L. D'autre part, les mutants pour le *TAS-X* ne sont disponibles seulement pour *Zucchini*. Des constructions de souches mutantes possédant le *TAS-X* seront nécessaires pour valider le rôle des autres protéines de la voie de biogenèse des piRNA pour tout les TAS-R.

Les piRNA du domaine F sont uniques

L'utilisation d'un transgène cible *pUASp-DomF-RFP-K10*, possédant la séquence du domaine F, a permis de démontrer que les piRNA issus du domaine F du *TAS-X* peuvent éteindre l'expression d'une séquence homologue. Ces piRNA sont donc fonctionnels. Une extinction de l'expression de la RFP aurait pu être possible par les piRNA des TAS-L (présents dans la souche *w¹¹¹⁸*) car il existe une séquence commune entre les TAS-L et le domaine F présentant une identité 80%, appelée la séquence *TLL* (Figure 47). Grâce à l'analyse des petits ARN les banques de données pour les souches *w¹¹¹⁸* et *P-1152w^c*, il a pu être montré que les piRNA produits à partir des deux séquences *TLL* sont différents. Une population de piRNA commune entre ces deux séquences n'apparaît que trois mésappariements sont utilisés pour l'alignement des piRNA sur la séquence du domaine F. Il n'existe donc aucun autre locus présent dans le génome produisant les mêmes piRNA que le domaine F. Par ailleurs, dans ces expériences, des piRNA homologues à l'élément *P* (apportés par la souche possédant le Gal4) sont aussi présents et participent aussi dans la répression de la cible. Afin de s'affranchir de cela, des souches recombinantes *pUASp-DomF-RFP-K10 ; Gal4* ont été construites. La présence du Gal4 dans la souche possédant le transgène permettra son expression de façon continue. Ainsi le génotype des souches apportant la source de piRNA correspondants au domaine F pourra être contrôlée plus efficacement.

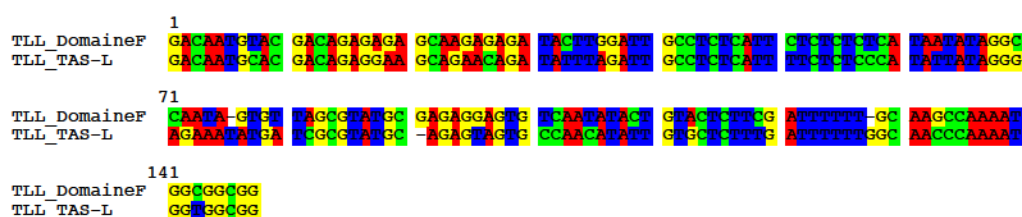


Figure 47. Alignement de la séquence *TLL* du domaine F et des TAS-L. Le programme MUSCLE a été utilisé pour l'alignement.

Rôle des piRNA de TAS dans la régulation de l'expression du génome

Un des projets de cette thèse était de rechercher si les piRNA de TAS sont impliqués dans la régulation de gènes en lignée germinale. Nous avons pu déterminer qu'il existe dans le génome de courtes séquences correspondants aux TAS ailleurs que dans les régions sub-téломériques. Ces séquences ont été retrouvées dans les transcrits des gènes ou bien dans les régions flanquantes de gènes. Une analyse de l'expression différentielle de l'expression de

gènes dans une souche ne possédant pas le *TAS-X* par rapport à une souche le possédant a été initiée au laboratoire. Nous avons malheureusement rencontrés de nombreux problèmes expérimentaux ayant ralenti l'obtention des données de séquençage adéquate du transcriptome. Une analyse préliminaire a permis de montrer qu'il existe 126 gènes sur exprimés en l'absence des piRNA issus du domaine F et possédant dans leur séquence un fragment complémentaire à ces piRNA. Ces gènes semblent indiquer un rôle des *TAS-X* dans la répression des gènes associés aux processus neuronaux. De futures expériences de RT-PCR quantitative doivent être effectuées afin de valider l'expression différentielle de ces gènes. Par ailleurs, certains de ces gènes pourront être étudiés plus en détail par exemple grâce à l'utilisation de constructions transgéniques où la séquence complémentaire aux piRNA du domaine F sera mutée. Cette mutation entraînera une absence de régulation si celle-ci existe.

Transmission des propriétés répressives par les TAS

Au laboratoire, il avait été montré précédemment que les capacités de répression d'une cible *P-lacZ* par un transgène *P-lacZ* inséré dans les TAS s'acquièrent au cours des générations quand celui-ci est transmis paternellement (Josse et al, 2007). Ce type de transmission permet de mimer l'arrivée d'une nouvelle séquence dans le génome. Ces expériences ont été refaites en analysant cette fois-ci la production de piRNA au cours des générations. Nous avons pu montrer que l'augmentation des capacités de répression est accompagnée d'une augmentation de la quantité de piRNA du transgène inséré dans les TAS. Notre hypothèse est que les TAS transmettent progressivement au transgène leurs propriétés répressives qui passeraient par le dépôt de marques répressives ainsi que la présence du complexe RDC. Ceci s'accompagne de l'effacement progressif de l'identité du transgène qui passe plus particulièrement par l'établissement d'une transcription non canonique bloquant ensuite l'épissage des transcrits. Des expériences de ChIP-qPCR devront être effectuées dans le futur afin de suivre l'état chromatinien (marque H3K9me3) du transgène ainsi que l'arrivée du RDC au fil des générations sur le transgène. Ceci permettra de valider notre hypothèse initiale.

En conclusion, nous avons pu montrer que les TAS-R forment un cluster producteur de piRNA assez original par rapport aux autres clusters présents dans le génome. D'une part à cause de leur localisation sub-télomérique, et d'autre part grâce à leurs propriétés répressives. En effet, les TAS sont des locus polymorphiques qui peuvent être perdus au fil des générations par des cassures terminales notamment quand les drosophiles sont maintenues dans les conditions de culture des laboratoires. De plus, le *TAS-X* possède un statut particulier par rapport aux autres TAS-R car il est retrouvé systématiquement dans les populations naturelles où il pourrait avoir un rôle différent de celui qu'il pourrait avoir dans les laboratoires. La domestication des drosophiles au laboratoire pourrait avoir un effet sur la régulation des gènes impliqués dans les processus neuronaux (Stanley and Kulathinal 2016). Une hypothèse est que les piRNA issus des locus TAS pourraient réguler l'expression de certains de ces gènes. Par ailleurs, les locus TAS sont un point chaud d'insertion de l'élément *P*. Ce locus transmet aux séquences qui s'y insèrent des propriétés répressives permettant l'établissement de leur répression. Ce locus qui pourrait être qualifié de « trappe génomique » ou de « piège à loup », assure donc un rôle important dans la protection des génomes contre les effets néfastes que pourraient provoquer l'arrivée d'un nouvel élément transposable.

IV-Matériels et Méthodes

A-Hybridation *in situ* sur chromosomes polytènes

L'hybridation *in situ* sur chromosomes polytènes de drosophiles est une méthode pour détecter et localiser la présence d'un fragment d'ADN de la séquence souhaitée. Les chromosomes polytènes sont issus de larves de troisième stade. C'est à ce moment du développement que des chromosomes avec un nombre important de cycle de réplication peuvent être obtenus ($n=800$). Par ailleurs il n'y a pas de séparations des chromatides dans ces chromosomes.

A1. Préparation des squashes et de la sonde ADN.

Obtention des chromosomes polytènes

Matériel : levure déshydratée, eau et tubes avec milieu nutritif.

Afin d'obtenir de jolis chromosomes polytènes, il faut mettre à pondre environ 5 femelles et 5 males à 18°C dans un tube préalablement stérilisé et dans lequel on a rajouté un peu de levure déshydratée. Il faut éviter de mettre trop d'individus dans un tube car cela risquerait de créer une surpopulation et pouvant conduire à un manque de nutriments pour le bon développement des larves. Le tube sera repiqué au moins deux fois tout les 2-3 jours. Il ne faut pas oublier de vider le tube après le dernier repiquage.

Il faut surveiller les tubes de temps en temps pour vérifier que le développement des larves se passe correctement. Au besoin, rajouter quelques gouttes d'eau si le milieu est trop sec. Des larves L3 au bout de 8 à 10 jours peuvent être obtenues selon les souches utilisées.

Squash

Matériel : acide acétique 45%, lames polysine (VWR 631-0107), lamelles 24x40mm, crayon à papier avec gomme et pinces à dissection.

La technique du squash permet d'étaler les chromosomes issus des glandes salivaires de larves de 3^{ème} stades sur une lame.

Pour cela, il faut disséquer les larves dans de l'acide acétique 45% sous la loupe binoculaire.

A l'aide de pinces, il faut séparer la tête du corps par étirement progressif afin de libérer et récupérer les glandes salivaires.

Ensuite, déposer les glandes salivaires sur une lame polysine dans une goutte d'acide acétique 45% et déposer délicatement une lamelle.

Commence maintenant le squash a proprement dit, à l'aide d'une gomme de crayon à papier, tapoter sur l'endroit la lamelle à l'endroit où les glandes salivaires ont été déposées. Plus l'on va tapoter et plus l'on va écraser les chromosomes des glandes salivaires. Il faut donc trouver le bon équilibre entre étalement des chromosomes et force exercée sur la lamelle. Cette étape peut être contrôlée et suivie à l'aide d'un microscope.

Fixation des chromosomes

Matériel : azote liquide, éthanol 90% et scalpel.

Une fois les chromosomes bien étalés sur la lame. Plonger celle-ci dans de l'azote liquide 5 pendant minutes. Puis retirer rapidement la lamelle à l'aide d'un scalpel et plonger la lame dans de l'éthanol 90% au minimum 10 minutes.

Après cette étape les chromosomes sont prêts à être utilisés pour faire des marquages. Ils peuvent être stockés et conservés plusieurs mois. Avant de procéder à l'hybridation *in situ*, un tri des chromosomes les mieux étalés doit être effectué (Figure 48).



Figure 48. Observation des chromosomes polytènes après les étapes de squashes et de fixation au microscope à contraste de phase.

J'ai utilisé par la suite deux techniques différentes pour faire des marquages sur chromosomes polytènes. La première nécessite l'utilisation du DAB, il s'agit d'une méthode cytochimique. La seconde est une méthode fluorescente utilisant des anticorps couplé à un fluorochrome.

Préparation des sondes ADN marquées

Matériel : BioNick™ DNA Labeling System (Life Technologies 18247-015) et DIG-Nick translation mix (Roche 11745816910)

Le marquage de l'ADN se fait par nick translation en utilisant 1µg d'ADN dans les deux cas. Toutes les instructions concernant les marquages sont disponibles chez les différents fournisseurs.

A2. Hybridation *in situ* avec sonde ADN biotinylée

Pré-hybridation

Matériel : SSC 2X, Ethanol 90%, NaOH 0.07N, 3 cuves, porte-lame

Mettre du SSC 2X à préchauffer à 65°C.

Plonger les lames pendant 30min dans le SSC 2X à 65°C

Retirer l'excédant de SSC et plonger les lames dans un bain d'éthanol à 90% pendant 5min

Retirer les lames et les plonger dans un nouveau bain d'éthanol 90% pendant 15min

Laisser sécher les lames.

Dénaturation

Matériel : Ethanol 90%, NaOH 0.07N, 2 cuves, porte-lame

Préparer du NaOH 0.07 extemporanément.

Plonger les lames 2 min dans le NaOH 0.07N. Attention à ne pas dépasser ce temps car cela pourrait abîmer les ADN.

Plonger les lames immédiatement dans un bain d'éthanol 90%.

Laisser sécher les lames.

Préparation de la solution d'hybridation

Matériel : ADN marqué au préalable (DIG ou biotine), sulfate de dextran 50%, formamide, SSC 20X, bain marie à 90°, glace

Dans un tube mettre : 10µL d'ADN marqué

10µL de sulfate de dextran 50%

10µL de SSC 20X

30µL de formamide

Homogénéiser la solution.

Dénaturer 8 min à 90°C

Mettre le tube dans la glace immédiatement.

Hybridation

Matériel : solution d'hybridation, lames dénaturées, lamelles 20x20mm, boîte hermétique, étuve 37°C.

Déposer 4µL de la solution d'hybridation sur les lames. Il faut essayer de déposer la goutte bien au centre du squash.

Mettre délicatement une lamelle

Placer les lames dans une chambre hermétique humidifiée avec du SSC 2X à 37°C toute la nuit.

Détection et révélation

Matériel : SSC 2X, PBS 1X, BSA 10%, H₂O, extravidine peroxydase (Sigma E2886), lamelles 24x40mm, DAB, H₂O₂ 1%, chambre noire, Giemsa, tampon phosphate et Eukitt

Laver les lames dans du SSC 2X deux fois 5min à température ambiante.

Les rincer dans du PBS 1X.

Préparer la solution de détection avec 100µL de BSA 10%, 90µL d'H₂O et 4µL d'extravidine peroxydase.

Déposer 100µL de cette solution sur les lames.

Incuber 30 min à 37°C en chambre humide.

Rincer les lames dans du PBS 1X

Préparer la solution de révélation avec un peu de DAB dans 500µL de PBS 1X. Rajouter 100µL H₂O₂ 1% juste avant l'emploi. A partir de ce moment, il faut travailler rapidement et surtout à l'obscurité car le DAB est photosensible. Il faut aussi faire attention du DAB car c'est un produit très toxique et sa manipulation devra être faite sous la hotte. Il ne faut pas hésiter aussi à changer de paire de gants lors de son utilisation afin d'éviter les contaminations.

Déposer 500µL de cette solution sur les lames et les placer à l'obscurité pendant 8min.

Pendant ce temps, préparer la solution de coloration. Il s'agit d'un tampon Giemsa. Pour cela, Diluer 3mL de Giemsa dans 94mL d'H₂O. Rajouter ensuite 3mL du tampon phosphate.

Après la révélation, rincer les lames dans du PBS 1X.

Les plonger dans le tampon Giemsa pendant 4min.

Rincer les lames à l'eau claire et les faire sécher.

Déposer une goutte d'Eukitt au centre de lame et la recouvrir d'une lamelle 24x40mm.

Laisser sécher la résine de montage environ 1h à température ambiante avant l'observation.

Les lames sont prêtes à être observées au microscope (Figure 49).

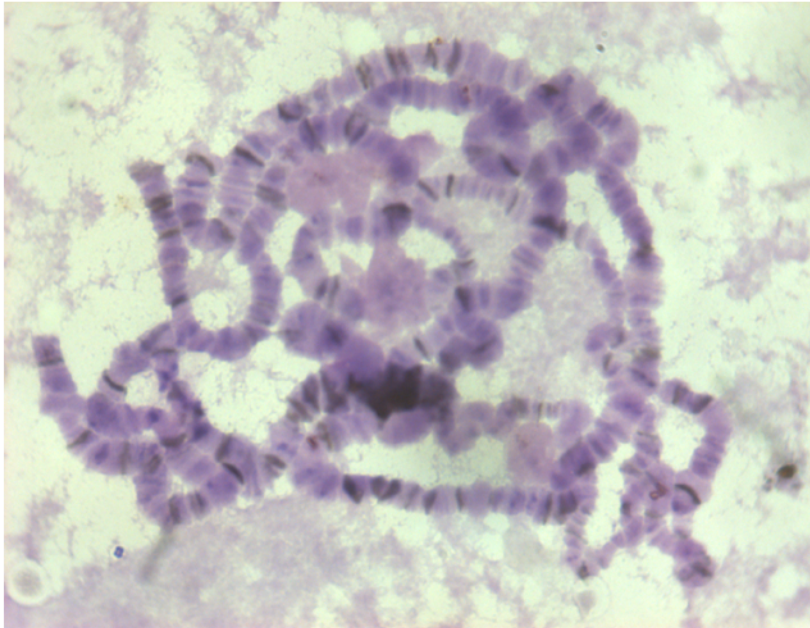


Figure 49. Observation d'un squash de chromosomes polytènes après hybridation *in situ* avec une sonde pour l'élément *P* marquée à la biotine et coloration au Giemsa.

A3. Hybridation *in situ* avec sonde ADN fluorescente

L'hybridation *in situ* avec des sondes fluorescentes m'a permis d'effectuer des expériences de co-localisation entre deux séquences d'ADN (l'élément *P* et les TAS). Pour cette technique, il faut donc marquer les sondes avec deux molécules différentes (DIG et biotine par exemple).

Les étapes de pré-hybridation et dénaturation sont les mêmes que celles décrites B1 et B2.

Préparation de la sonde

Matériel : ADN marqué au préalable (DIG et biotine), sulfate de dextran 50%, formamide, SSC 20X, bain marie à 90°, glace

Dans un tube mettre : 10µL d'ADN marqué à la DIG

10µL d'ADN marqué à la biotine

10µL de sulfate de dextran 50%

10µL de SSC 20X

30µL de formamide

Homogénéiser la solution.

Dénaturer 8 min à 90°C

Mettre le tube dans la glace immédiatement.

Hybridation

Matériel : solution d'hybridation, lames dénaturées, lamelles 20x20mm, boîte hermétique, étuve 37°C.

Déposer 10µL de la solution d'hybridation sur les lames. Il faut essayer de déposer la goutte bien au centre du squash.

Mettre délicatement une lamelle

Placer les lames dans une chambre hermétique humidifiée avec du SSC 2X à 37°C toute la nuit.

Détection et révélation

Matériel : SSC 2X, solution STB1, solution STB5*, lamelles 24x40mm, anti-DIG rhodamine (Roche 11207750910), extravidine -FITC (Sigma E2761), tampon de lavage, Vectashield® (H-1000), DAPI.*

Mettre du SSC 2X ainsi que le tampon de lavage à chauffer à 37°C au préalable

Laver les lames deux fois 10 min dans du SSC 2X à 37°C

Retirer l'excédant de SSC 2X

Mettre 80µL de solution STB5 au centre du squash et recouvrir d'une lamelle.

Incuber 30 min à 37°C

Retirer l'excédant de la solution STB5 en faisant délicatement glisser la lamelle

Mettre 80µL de solution STB1 au centre du squash et recouvrir d'une lamelle

Incuber 30 min à 37°C

Laver les lames deux fois 5 min dans le tampon de lavage à 37°C

Faire sécher 2-3 min afin de retirer l'excès de tampon de lavage

Mettre 15µL de Vectashield® auquel le DAPI a été rajouté au préalable.

Les lames sont prêtes pour l'observation du microscope à fluorescence (Figure 50).

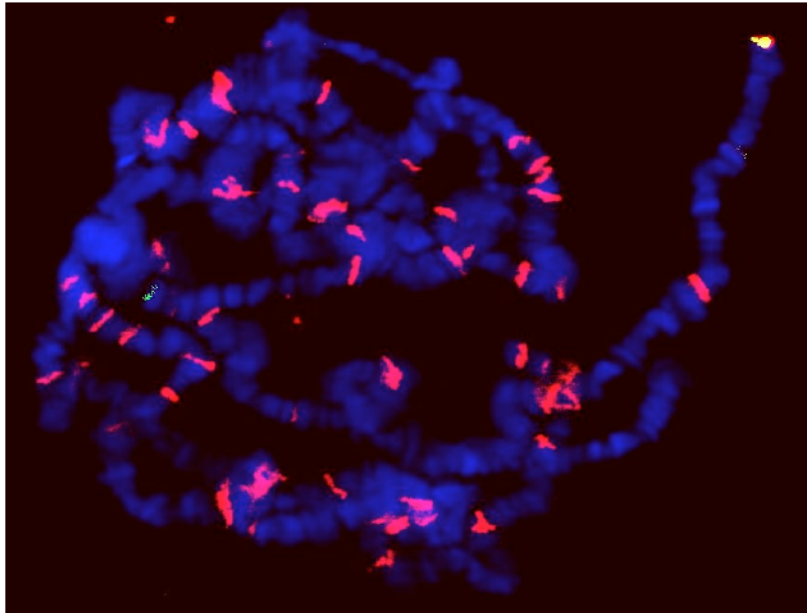


Figure 50. Observation d'un squash de chromosomes polytènes après hybridation *in situ* avec une sonde pour l'élément *P* marquée à la DIG (rouge) et une sonde TAS marquée à la biotine (vert).

Préparation des solutions

-NaOH 0.07N

Peser 2,8g de cristaux de NaOH pour 1L d'H₂O

-Sulfate de Dextran 50%

Peser 50mg de sulfate de Dextran pour 100µL d'H₂O

-Tampon phosphate

Peser 7,39g de KH₂PO₄ rajouter 800mL d'H₂O

Ajuster le pH à 6.7 avec du Na₂HPO₄.2H₂O

Compléter pour un volume final de 1 litre et conserver à 4°C

-STB1

Mélanger dans l'ordre : 200µL SSC 20X, 800µL d'H₂O, 10mg de BSA et 1µL de triton 100X

-STB5

Mélanger dans l'ordre : 200µL SSC 20X, 800µL d'H₂O, 50mg de BSA et 1µL de triton 100X

-Tampon de lavage

Mélanger dans l'ordre : 40mL de SSC 2X, 160mL d'H₂O et 200µL de triton 100X

Remarques

-La solution de dextran est très visqueuse. Elle doit être stockée à 4°C et renouvelée assez fréquemment (toute les semaines).

-Toutes les solutions (SSC 2X, PBS 1X, Ethanol 90%) doivent être fraîchement préparées pour chaque expérience et peuvent être conservées quelques jours.

-Il faut faire des aliquotes pour l'extravidine-FITC ainsi que l'anti-DIG Rhodamine. Les conservés à -20°C.

-La solution d'anti-DIG rhodamine en cours d'utilisation doit être gardée à 4°C. Ne pas hésiter à prendre un nouvel aliquote quand de l'hybridation aspécifique (cristaux rouges dispersés) apparaît sur les lames.

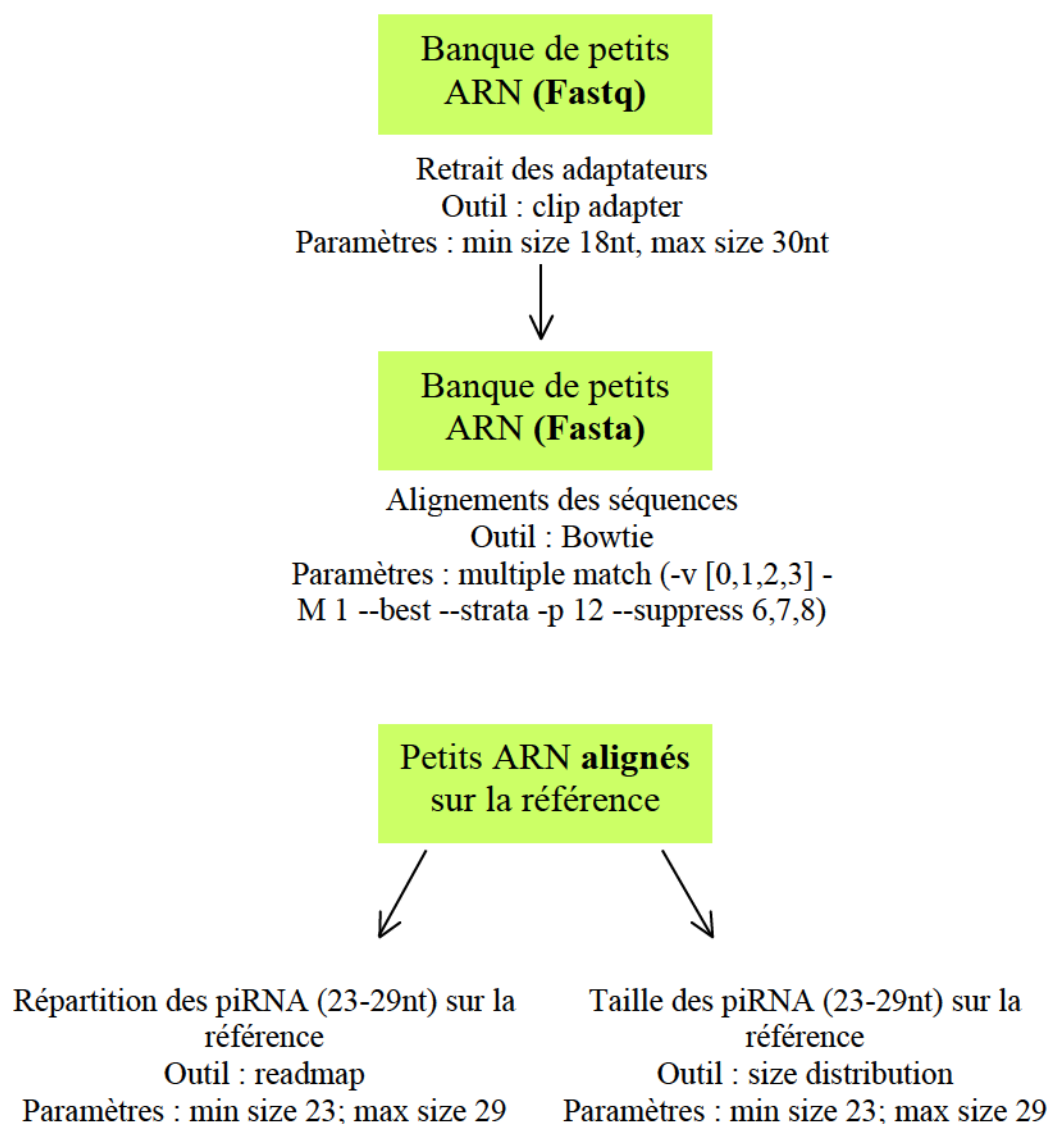
-Faire attention lors de l'utilisation du DAB car c'est un produit toxique. Il faut le manipuler sous la hotte

-Les lames préparées à l'aide de la méthode cytochimique peuvent être conservées très longtemps à température ambiante à l'abri.

-Les lames préparées à l'aide de la méthode fluorescente peuvent être conservées à 4°C quelques mois.

B. Analyse des petits ARN

Au laboratoire, nous disposons d'un accès à « Galaxy » regroupant un grand nombre d'outils informatiques nécessaires pour l'analyse de données de séquençage à haut débit. Cet accès est géré par la plate-forme bio-informatique (ARTbio) de l'Institut de Biologie Paris-Seine. Les différentes étapes pour l'analyse sont présentées dans la figure .



C. Détection de la fluorescence endogène en ovaires de drosophile

Voici le protocole expérimental suivi pour analyser l'expression endogène de la RFP (Red Fluorescent Protein) lors de l'étude des transgènes *pUASp-DomF-RFP-K10*.

Disséquer les ovaires sous la loupe binoculaire dans du PBS 1X.

Fixation des ovaires dans du PBS-formaldéhyde 5% pendant 15 min à température ambiante.

Faire trois lavages dans du PBS 1X pendant 15 min à température ambiante.

Dans le dernier lavage, rajouter du DAPI dilué au 1/1000 dans du PBS 1X pendant 20 min à l'obscurité.

Faire le montage des ovaires entre lames et lamelles dans une goutte de PBS-glycérol.

Pendant cette étape, il se faut débarrasser de tous les tissus superflus tels que les tubes de Malpighi ou des fragments de cuticules pouvant être restés lors de la dissection. Les ovaires sont assez fragiles malgré la fixation, il faut essayer d'écarter au maximum les ovarioles sans détruire leur structure.

Observer au microscope confocal.

V-Références bibliographiques

Abad, J. P., B. De Pablos, K. Osoegawa, P. J. De Jong, A. Martin-Gallardo and A. Villasante (2004a). "TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres." Mol Biol Evol **21**(9): 1620-1624.

Abad, J. P., B. de Pablos, M. Agudo, I. Molina, G. Giovinazzo, A. Martin-Gallardo and A. Villasante (2004b). "Genomic and cytological analysis of the Y chromosome of *Drosophila melanogaster*: telomere-derived sequences at internal regions." Chromosoma **113**(6): 295-304.

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, . . . J. C. Venter (2000). "The genome sequence of *Drosophila melanogaster*." Science **287**(5461): 2185-2195.

Andreyeva, E. N., E. S. Belyaeva, V. F. Semeshin, G. V. Pokholkova and I. F. Zhimulev (2005). "Three distinct chromatin domains in telomere ends of polytene chromosomes in *Drosophila melanogaster* Tel mutants." J Cell Sci **118**(Pt 23): 5465-5477.

Antao, J. M., J. M. Mason, J. Dejardin and R. E. Kingston (2012). "Protein landscape at *Drosophila melanogaster* telomere-associated sequence repeats." Mol Cell Biol **32**(12): 2170-2182.

Ayarpadikannan, S. and H. S. Kim (2014). "The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases." Genomics Inform **12**(3): 98-104.

Barckmann, B., S. Pierson, J. Dufourt, C. Papin, C. Armenise, F. Port, . . . M. Simonelig (2015). "Aubergine iCLIP Reveals piRNA-Dependent Decay of mRNAs Involved in Germ Cell Development in the Early Embryo." Cell Rep **12**(7): 1205-1216.

Bergman, C. M., H. Quesneville, D. Anxolabehere and M. Ashburner (2006). "Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome." Genome Biol **7**(11): R112.

Biemont, C. (2010). "A brief history of the status of transposable elements: from junk DNA to major players in evolution." Genetics **186**(4): 1085-1093.

Biemont, C. and C. Vieira (2006). "Genetics: junk DNA as an evolutionary force." Nature **443**(7111): 521-524.

Boivin, A., C. Gally, S. Netter, D. Anxolabehere and S. Ronsseray (2003). "Telomeric associated sequences of *Drosophila* recruit polycomb-group proteins in vivo and can induce pairing-sensitive repression." Genetics **164**(1): 195-208.

Brennecke, J., C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark and G. J. Hannon (2008). "An epigenetic role for maternally inherited piRNAs in transposon silencing." Science **322**(5906): 1387-1392.

Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam and G. J. Hannon (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*." Cell **128**(6): 1089-1103.

Cordaux, R. and M. A. Batzer (2009). "The impact of retrotransposons on human genome evolution." Nat Rev Genet **10**(10): 691-703.

Czech, B., J. B. Preall, J. McGinn and G. J. Hannon (2013). "A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway." Mol Cell **50**(5): 749-761.

de Castro, J. P. and C. M. Carareto (2004). "Canonical P elements are transcriptionally active in the saltans group of *Drosophila*." J Mol Evol **59**(1): 31-40.

de Vanssay, A., A. L. Bouge, A. Boivin, C. Hermant, L. Teyssset, V. Delmarre, . . . S. Ronsseray (2012). "Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus." Nature **490**(7418): 112-115.

Donertas, D., G. Sienski and J. Brennecke (2013). "*Drosophila* Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex." Genes Dev **27**(15): 1693-1705.

Doolittle, W. F. and C. Sapienza (1980). "Selfish genes, the phenotype paradigm and genome evolution." Nature **284**(5757): 601-603.

Dorer, D. R. and S. Henikoff (1994). "Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*." Cell **77**(7): 993-1002.

Dorer, D. R. and S. Henikoff (1997). "Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans." Genetics **147**(3): 1181-1190.

Eddy, E. M. (1974). "Fine structural observations on the form and distribution of nuage in germ cells of the rat." Anat Rec **178**(4): 731-757.

Engels, W. R. (1979). "Germ line aberrations associated with a case of hybrid dysgenesis in *Drosophila melanogaster* males." Genet Res **33**(2): 137-146.

Feschotte, C. and E. J. Pritham (2007). "DNA transposons and the evolution of eukaryotic genomes." Annu Rev Genet **41**: 331-368.

Font-Burgada, J., D. Rossell, H. Auer and F. Azorin (2008). "*Drosophila* HP1c isoform interacts with the zinc-finger proteins WOC and Relative-of-WOC to regulate gene expression." Genes Dev **22**(21): 3007-3023.

Frydrychova, R. C., J. M. Mason and T. K. Archer (2008). "HP1 is distributed within distinct chromatin domains at *Drosophila* telomeres." Genetics **180**(1): 121-131.

Fulcher, N., E. Derboven, S. Valuchova and K. Riha (2014). "If the cap fits, wear it: an overview of telomeric structures over evolution." Cell Mol Life Sci **71**(5): 847-865.

George, J. A., P. G. DeBaryshe, K. L. Traverse, S. E. Celniker and M. L. Pardue (2006). "Genomic organization of the *Drosophila* telomere retrotransposable elements." Genome Res **16**(10): 1231-1240.

Goriaux, C., S. Desset, Y. Renaud, C. Vaury and E. Brasset (2014). "Transcriptional properties and splicing of the flamenco piRNA cluster." EMBO Rep **15**(4): 411-418.

Gou, L. T., P. Dai, J. H. Yang, Y. Xue, Y. P. Hu, Y. Zhou, . . . M. F. Liu (2014). "Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis." Cell Res **24**(6): 680-700.

Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, . . . M. C. Siomi (2007). "A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*." Science **315**(5818): 1587-1590.

Haase, A. D., S. Fenoglio, F. Muerdter, P. M. Guzzardo, B. Czech, D. J. Pappin, . . . G. J. Hannon (2010). "Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*." Genes Dev **24**(22): 2499-2504.

Han, B. W., W. Wang, C. Li, Z. Weng and P. D. Zamore (2015). "Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production." Science **348**(6236): 817-821.

Handler, D., K. Meixner, M. Pizka, K. Lauss, C. Schmied, F. S. Gruber and J. Brennecke (2013). "The genetic makeup of the *Drosophila* piRNA pathway." Mol Cell **50**(5): 762-777.

He, L., X. Wang and D. J. Montell (2011). "Shining light on *Drosophila* oogenesis: live imaging of egg development." Curr Opin Genet Dev **21**(5): 612-619.

Hermant, C., A. Boivin, L. Teyssset, V. Delmarre, A. Asif-Laidin, M. van den Beek, . . . S. Ronsseray (2015). "Paramutation in *Drosophila* Requires Both Nuclear and Cytoplasmic Actors of the piRNA Pathway and Induces Cis-spreading of piRNA Production." Genetics **201**(4): 1381-1396.

Hiraizumi, Y. and J. F. Crow (1960). "Heterozygous Effects on Viability, Fertility, Rate of Development, and Longevity of *Drosophila* Chromosomes That Are Lethal When Homozygous." Genetics **45**(8): 1071-1083.

Hirakata, S. and M. C. Siomi (2016). "piRNA biogenesis in the germline: From transcription of piRNA genomic sources to piRNA maturation." Biochim Biophys Acta **1859**(1): 82-92.

Huang, X. A., H. Yin, S. Sweeney, D. Raha, M. Snyder and H. Lin (2013). "A major epigenetic programming mechanism guided by piRNAs." Dev Cell **24**(5): 502-516.

Ishizu, H., Y. W. Iwasaki, S. Hirakata, H. Ozaki, W. Iwasaki, H. Siomi and M. C. Siomi (2015). "Somatic Primary piRNA Biogenesis Driven by cis-Acting RNA Elements and trans-Acting Yb." Cell Rep **12**(3): 429-440.

Iwasaki, Y. W., M. C. Siomi and H. Siomi (2015). "PIWI-Interacting RNA: Its Biogenesis and Functions." Annu Rev Biochem **84**: 405-433.

Josse, T., L. Teyssset, A. L. Todeschini, C. M. Sidor, D. Anxolabehere and S. Ronsseray (2007). "Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation." PLoS Genet **3**(9): 1633-1643.

Josse, T., C. Maurel-Zaffran, A. de Vanssay, L. Teyssset, A. L. Todeschini, V. Delmarre, . . . S. Ronsseray (2008). "Telomeric trans-silencing in *Drosophila melanogaster*: tissue specificity, development and functional interactions between non-homologous telomeres." PLoS One **3**(9): e3249.

Karpen, G. H. and A. C. Spradling (1992). "Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome Dp1187 by single P element insertional mutagenesis." Genetics **132**(3): 737-753.

Kidwell, M. G., J. F. Kidwell and J. A. Sved (1977). "Hybrid Dysgenesis in *DROSOPHILA MELANOGASTER*: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination." Genetics **86**(4): 813-833.

Kim, A., C. Terzian, P. Santamaria, A. Pelisson, N. Purd'homme and A. Bucheton (1994). "Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*." Proc Natl Acad Sci U S A **91**(4): 1285-1289.

Kim, M., N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedeja, J. F. Greenblatt and S. Buratowski (2004). "The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II." Nature **432**(7016): 517-522.

Kiuchi, T., H. Koga, M. Kawamoto, K. Shoji, H. Sakai, Y. Arai, . . . S. Katsuma (2014). "A single female-specific piRNA is the primary determiner of sex in the silkworm." Nature **509**(7502): 633-636.

Klattenhoff, C., H. Xi, C. Li, S. Lee, J. Xu, J. S. Khurana, . . . W. E. Theurkauf (2009). "The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters." Cell **138**(6): 1137-1149.

Le Thomas, A., A. K. Rogers, A. Webster, G. K. Marinov, S. E. Liao, E. M. Perkins, . . . K. F. Toth (2013). "Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state." Genes Dev **27**(4): 390-399.

Le Thomas, A., E. Stuwe, S. Li, J. Du, G. Marinov, N. Rozhkov, . . . A. A. Aravin (2014). "Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing." Genes Dev **28**(15): 1667-1680.

Lehmann, R. and C. Nusslein-Volhard (1991). "The maternal gene nanos has a central role in posterior pattern formation of the *Drosophila* embryo." Development **112**(3): 679-691.

Levis, R. W., R. Ganesan, K. Houtchens, L. A. Tolar and F. M. Sheen (1993). "Transposons in place of telomeric repeats at a *Drosophila* telomere." Cell **75**(6): 1083-1093.

Malone, C. D., J. Brennecke, M. Dus, A. Stark, W. R. McCombie, R. Sachidanandam and G. J. Hannon (2009). "Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary." Cell **137**(3): 522-535.

Marin, L., M. Lehmann, D. Nouaud, H. Izaabel, D. Anxolabehere and S. Ronsseray (2000). "P-Element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric P copy." Genetics **155**(4): 1841-1854.

Mason, J. M. and A. Villasante (2014). Subtelomeres in *Drosophila* and Other Diptera. Subtelomeres. J. E. Louis and M. M. Becker. Berlin, Heidelberg, Springer Berlin Heidelberg: 211-225.

Mason, J. M., R. C. Frydrychova and H. Biessmann (2008). "*Drosophila* telomeres: an exception providing new insights." Bioessays **30**(1): 25-37.

Mevel-Ninio, M., A. Pelisson, J. Kinder, A. R. Campos and A. Bucheton (2007). "The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis." Genetics **175**(4): 1615-1624.

Misra, S. and D. C. Rio (1990). "Cytotype control of *Drosophila* P element transposition: the 66 kd protein is a repressor of transposase activity." Cell **62**(2): 269-284.

Moazed, D. (2012). "A piRNA to remember." Cell **149**(3): 512-514.

Mohn, F., D. Handler and J. Brennecke (2015). "Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis." Science **348**(6236): 812-817.

Mohn, F., G. Sienski, D. Handler and J. Brennecke (2014). "The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*." Cell **157**(6): 1364-1379.

Muerdter, F., I. Olovnikov, A. Molaro, N. V. Rozhkov, B. Czech, A. Gordon, . . . A. A. Aravin (2012). "Production of artificial piRNAs in flies and mice." RNA **18**(1): 42-52.

Murota, Y., H. Ishizu, S. Nakagawa, Y. W. Iwasaki, S. Shibata, M. K. Kamatani, . . . M. C. Siomi (2014). "Yb integrates piRNA intermediates and processing factors into perinuclear bodies to enhance piRISC assembly." Cell Rep **8**(1): 103-113.

Nishimasu, H., H. Ishizu, K. Saito, S. Fukuhara, M. K. Kamatani, L. Bonnefond, . . . O. Nureki (2012). "Structure and function of Zucchini endoribonuclease in piRNA biogenesis." Nature **491**(7423): 284-287.

O'Hare, K. and G. M. Rubin (1983). "Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome." Cell **34**(1): 25-35.

Ohtani, H., Y. W. Iwasaki, A. Shibuya, H. Siomi, M. C. Siomi and K. Saito (2013). "DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the Drosophila ovary." Genes Dev **27**(15): 1656-1661.

Olovnikov, I., S. Ryazansky, S. Shpiz, S. Lavrov, Y. Abramov, C. Vaury, . . . A. Kalmykova (2013). "De novo piRNA cluster formation in the Drosophila germ line triggered by transgenes containing a transcribed transposon fragment." Nucleic Acids Res **41**(11): 5757-5768.

Orgel, L. E. and F. H. Crick (1980). "Selfish DNA: the ultimate parasite." Nature **284**(5757): 604-607.

Pane, A., K. Wehr and T. Schupbach (2007). "zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline." Dev Cell **12**(6): 851-862.

Papin, C. and M. Simonelig (2011). "Contrôle du développement embryonnaire par des petits ARN issus de transposons." Med Sci (Paris) **27**(12): 1050-1052.

Pardue, M. L. and P. G. DeBaryshe (2008). "Drosophila telomeres: A variation on the telomerase theme." Fly (Austin) **2**(3): 101-110.

Pek, J. W., V. S. Patil and T. Kai (2012). "piRNA pathway and the potential processing site, the nuage, in the Drosophila germline." Dev Growth Differ **54**(1): 66-77.

Pelisson, A., S. U. Song, N. Prud'homme, P. A. Smith, A. Bucheton and V. G. Corces (1994). "Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene." EMBO J **13**(18): 4401-4411.

Perrat, P. N., S. DasGupta, J. Wang, W. Theurkauf, Z. Weng, M. Rosbash and S. Waddell (2013). "Transposition-driven genomic heterogeneity in the Drosophila brain." Science **340**(6128): 91-95.

Poyhonen, M., A. de Vanssay, V. Delmarre, C. Hermant, A. L. Todeschini, L. Teyssset and S. Ronsseray (2012). "Homology-dependent silencing by an exogenous sequence in the Drosophila germline." G3 (Bethesda) **2**(3): 331-338.

Raffa, G. D., L. Ciapponi, G. Cenci and M. Gatti (2011). "Terminin: a protein complex that mediates epigenetic maintenance of Drosophila telomeres." Nucleus **2**(5): 383-391.

Raffa, G. D., G. Cenci, L. Ciapponi and M. Gatti (2013). "Organization and Evolution of Drosophila Terminin: Similarities and Differences between Drosophila and Human Telomeres." Front Oncol **3**: 112.

Rajasethupathy, P., I. Antonov, R. Sheridan, S. Frey, C. Sander, T. Tuschl and E. R. Kandel (2012). "A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity." Cell **149**(3): 693-707.

Rangan, P., C. D. Malone, C. Navarro, S. P. Newbold, P. S. Hayes, R. Sachidanandam, . . . R. Lehmann (2011). "piRNA production requires heterochromatin formation in *Drosophila*." Curr Biol **21**(16): 1373-1379.

Rebollo, R., M. T. Romanish and D. L. Mager (2012). "Transposable elements: an abundant and natural source of regulatory sequences for host genes." Annu Rev Genet **46**: 21-42.

Reiss, D., T. Josse, D. Anxolabehere and S. Ronsseray (2004). "aubergine mutations in *Drosophila melanogaster* impair P cytotype determination by telomeric P elements inserted in heterochromatin." Mol Genet Genomics **272**(3): 336-343.

Robine, N., N. C. Lau, S. Balla, Z. Jin, K. Okamura, S. Kuramochi-Miyagawa, . . . E. C. Lai (2009). "A broadly conserved pathway generates 3'UTR-directed primary piRNAs." Curr Biol **19**(24): 2066-2076.

Roche, S. E. and D. C. Rio (1998). "Trans-silencing by P elements inserted in subtelomeric heterochromatin involves the *Drosophila* Polycomb group gene, Enhancer of zeste." Genetics **149**(4): 1839-1855.

Ronsseray, S., M. Lehmann and D. Anxolabehere (1989). "Copy number and distribution of P and I mobile elements in *Drosophila melanogaster* populations." Chromosoma **98**(3): 207-214.

Ronsseray, S., M. Lehmann and D. Anxolabehere (1991). "The maternally inherited regulation of P elements in *Drosophila melanogaster* can be elicited by two P copies at cytological site 1A on the X chromosome." Genetics **129**(2): 501-512.

Ronsseray, S., A. Boivin and D. Anxolabehere (2001). "P-Element repression in *Drosophila melanogaster* by variegating clusters of P-lacZ-white transgenes." Genetics **159**(4): 1631-1642.

Ronsseray, S., M. Lehmann, D. Nouaud and D. Anxolabehere (1996). "The regulatory properties of autonomous subtelomeric P elements are sensitive to a Suppressor of variegation in *Drosophila melanogaster*." Genetics **143**(4): 1663-1674.

Ronsseray, S., M. Lehmann, D. Nouaud and D. Anxolabehere (1997). "P element regulation and X-chromosome subtelomeric heterochromatin in *Drosophila melanogaster*." Genetica **100**(1-3): 95-107.

Rouget, C., C. Papin, A. Boureux, A. C. Meunier, B. Franco, N. Robine, . . . M. Simonelig (2010). "Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo." Nature **467**(7319): 1128-1132.

Saito, K., H. Ishizu, M. Komai, H. Kotani, Y. Kawamura, K. M. Nishida, . . . M. C. Siomi (2010). "Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*." Genes Dev **24**(22): 2493-2498.

- Saito, K., S. Inagaki, T. Mituyama, Y. Kawamura, Y. Ono, E. Sakota, . . . M. C. Siomi (2009). "A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*." Nature **461**(7268): 1296-1299.
- Sasaki, T. and H. Fujiwara (2000). "Detection and distribution patterns of telomerase activity in insects." Eur J Biochem **267**(10): 3025-3031.
- Schupbach, T. and E. Wieschaus (1991). "Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology." Genetics **129**(4): 1119-1136.
- Sienski, G., D. Donertas and J. Brennecke (2012). "Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression." Cell **151**(5): 964-980.
- Sienski, G., J. Batki, K. A. Senti, D. Donertas, L. Tirian, K. Meixner and J. Brennecke (2015). "Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery." Genes Dev **29**(21): 2258-2271.
- Siomi, M. C., K. Sato, D. Pezic and A. A. Aravin (2011). "PIWI-interacting small RNAs: the vanguard of genome defence." Nat Rev Mol Cell Biol **12**(4): 246-258.
- Smith, C. D., S. Shu, C. J. Mungall and G. H. Karpen (2007). "The Release 5.1 annotation of *Drosophila melanogaster* heterochromatin." Science **316**(5831): 1586-1591.
- Stanley, C. E., Jr. and R. J. Kulathinal (2016). "Genomic signatures of domestication on neurogenetic genes in *Drosophila melanogaster*." BMC Evol Biol **16**(1): 6.
- Teyssset, L., J. C. Burns, H. Shike, B. L. Sullivan, A. Bucheton and C. Terzian (1998). "A Moloney murine leukemia virus-based retroviral vector pseudotyped by the insect retroviral gypsy envelope can infect *Drosophila* cells." J Virol **72**(1): 853-856.
- Todeschini, A. L., L. Teyssset, V. Delmarre and S. Ronsseray (2010). "The epigenetic trans-silencing effect in *Drosophila* involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1." PLoS One **5**(6): e11032.
- Traverse, K. L. and M. L. Pardue (1988). "A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres." Proc Natl Acad Sci U S A **85**(21): 8116-8120.
- Vagin, V. V., Y. Yu, A. Jankowska, Y. Luo, K. A. Wasik, C. D. Malone, . . . G. J. Hannon (2013). "Minotaur is critical for primary piRNA biogenesis." RNA **19**(8): 1064-1077.
- Walter, M. F., C. Jang, B. Kasravi, J. Donath, B. M. Mechler, J. M. Mason and H. Biessmann (1995). "DNA organization and polymorphism of a wild-type *Drosophila* telomere region." Chromosoma **104**(4): 229-241.

- Wang, S. H., R. Nan, M. C. Accardo, M. Sentmanat, P. Dimitri and S. C. Elgin (2014). "A distinct type of heterochromatin at the telomeric region of the *Drosophila melanogaster* Y chromosome." PLoS One **9**(1): e86451.
- Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, . . . A. H. Schulman (2007). "A unified classification system for eukaryotic transposable elements." Nat Rev Genet **8**(12): 973-982.
- Yamanaka, S., M. C. Siomi and H. Siomi (2014). "piRNA clusters and open chromatin structure." Mob DNA **5**: 22.
- Yin, H. and H. Lin (2007). "An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*." Nature **450**(7167): 304-308.
- Yu, Y., J. Gu, Y. Jin, Y. Luo, J. B. Preall, J. Ma, . . . G. J. Hannon (2015). "Panoramix enforces piRNA-dependent cotranscriptional silencing." Science **350**(6258): 339-342.
- Zanni, V., A. Eymery, M. Coiffet, M. Zytnicki, I. Luyten, H. Quesneville, . . . S. Jensen (2013). "Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters." Proc Natl Acad Sci U S A **110**(49): 19842-19847.
- Zhang, L. and Y. S. Rong (2012). "Retrotransposons at *Drosophila* telomeres: host domestication of a selfish element for the maintenance of genome integrity." Biochim Biophys Acta **1819**(7): 771-775.
- Zhang, Z., J. Wang, N. Schultz, F. Zhang, S. S. Parhad, S. Tu, . . . W. E. Theurkauf (2014). "The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing." Cell **157**(6): 1353-1363.

VI-Annexes

Gènes avec piRNA du domaine F sur les deux brins transcrits	Gènes avec piRNA du domaine F sur le brin du transcrit	Gènes avec piRNA du domaine F sur le brin complémentaire du transcrit
AdamTS-A	alpha-Est10	amd
aux	Axn	Atg1
bcd	babo	bif
bip1	Baldspot	c(3)G
bol	bbx	C3G
br	beta4GalNAcTA	Cad86C
BTBD9	BicC	Cad88C
btsz	blot	CAH2
Cad87A	boi	CASK
Cadps	Br140	Cf2
CenG1A	Ca-beta	CG10013
CG10321	Cad74A	CG10407
CG10426	Cdc7	CG10492
CG10738	CG10011	CG10654
CG12299	CG10445	CG11034
CG13375	CG11147	CG11122
CG13933	CG11399	CG12179
CG14073	CG12194	CG12708
CG14629	CG12708	CG12990
CG14879	CG12769	CG15890
CG1632	CG12773	CG17167
CG1673	CG12877	CG18135
CG17169	CG1358	CG2658
CG17265	CG14309	CG2918
CG2258	CG15570	CG31342
CG2841	CG16758	CG31869
CG30438	CG17544	CG33111
CG31324	CG2187	CG34348
CG32152	CG2218	CG34449
CG32187	CG2941	CG3655
CG3281	CG3036	CG3800
CG3588	CG30428	CG3857
CG3699	CG31522	CG40298
CG3764	CG31689	CG42327
CG3777	CG32000	CG42450
CG42389	CG32572	CG42668
CG42797	CG32645	CG42681
CG4281	CG34166	CG42699
CG43313	CG4119	CG43333
CG4360	CG4198	CG44245
CG4433	CG4266	CG44838
CG5065	CG45084	CG45068
CG5151	CG5921	CG45069
CG5191	CG5966	CG4733
CG5270	CG5973	CG5273
CG5337	CG6495	CG5644
CG5414	CG7185	CG6805

CG5742	CG7365	CG8245
CG6752	CG7627	CG9747
CG7565	CG8713	Cht3
CG8405	CG9384	ck
CG8516	CG9914	corn
CG8929	CG9977	corolla
CG9380	CG9988	CR44462
CG9990	chrb	CR44779
CirI	clu	CR45046
Cpsf160	comt	D2hgdh
CR43144	CR44124	Debcl
CR43481	crm	Dhit
CR43836	CRMP	Dlg5
CrebA	cue	dop
dati	DAAM	Ect4
dgo	Dab	Efa6
Dh44-R2	Dark	Efr
DI	Dhc98D	egh
dpr6	didum	eIF5B
dsb	e(y)3	Eip74EF
dsd	EndoA	ena
E23	Fatp	fog
eIF-4B	Fkbp14	foi
ewg	flam	fs(1)M3
FASN3	fs(1)h	fs(1)N
Fcp3C	fz3	fus
fuss	GalNAc-T1	G9a
gb	GalNAc-T2	garz
GlcT-1	Gk2	Gfat1
gukh	GlcAT-P	Ggt-1
Hip14	glec	GLaz
Ir75a	Gprk2	GlcAT-P
JIL-1	HDAC4	Glycogenin
jvl	hep	HDAC6
kek6	kuk	Hers
ken	Liprin-alpha	Invadolysin
Lar	ImgA	jbug
mbt	ImgB	l(3)02640
Mer	LRP1	LBR
mGluR	Lsp2	Lsd-2
milt	mirr	mei-217
modSP	NaPi-III	mei-218
MRP	Nhe2	MESR4
mRpL12	niki	mia
mtt	norpA	mid
neo	numb	mnb
NUCB1	Obp19c	otu
ocm	Ork1	oxl
olf413	osk	Pde6
otu	pad	Peritrophin-A

Papss	path	PH4alphaEFB
Pde9	pigeon	Pp2C1
Pdk1	Plc21C	ppk29
pgant6	pncr014:3L	psidin
PI4KIIIalpha	psd	Pten
PICK1	Pzl	retn
Pkc98E	Rab27	Rox8
Pld	Rbcn-3A	RtGEF
psq	RhoGAP19D	Samuel
pum	RhoGAP1A	Shark
qkr58E-2	RhoGEF64C	sina
rut	rin	slam
scaf	robo1	slpr
SCAP	Sans	Spps
Sec16	sdk	Spt6
Set2	simj	su(w[a])
stet	SKIP	subdued
tamo	snz	swm
Tao	ssh	tal-1A
Tlk	Trpm	tal-2A
Topors	tty	tal-3A
ttk	Ubr3	tal-AA
tyn	veil	tay
Vm34Ca	Vm26Aa	tko
Vrp1	Vps26	Tollo
wapl	waw	topi
wge	yata	trx
Wnt5	yellow-g	Usf
yin	yrt	usp
		Vm26Ab
		Vml
		Wnk
		woc
		yellow-k
		z

Table 8. Liste des gènes dérégulés possédant une séquence de piRNA du domaine F

Structure et fonction des séquences subtélomériques productrices de piRNA

Les TAS (Telomeric Associated Sequences) sont des régions subtélomériques répétées non codantes formant un locus hétérochromatique chez *Drosophila melanogaster*. Il existe deux grandes familles de TAS, les TAS-R et les TAS-L possédant une structure et des propriétés différentes. Durant cette thèse, j'ai montré que les TAS dériveraient d'une séquence commune appelée *TLL* rapprochant ces deux familles. Par ailleurs, une étude des populations de drosophiles récoltées récemment dans la nature a permis de montrer qu'il existe d'une part une pression de sélection pour la présence du *TAS-X* dans la nature, alors que celui peut être perdu quand les drosophiles sont maintenues dans les laboratoires pendant plusieurs générations. Le *TAS-X* jouerait un rôle important dans la nature où il pourrait réguler l'expression de gènes par les piRNA. Et d'autre part, il n'existe aucune souche dépourvue de TAS-R à tout les télomère. Deux locus TAS-R ont toujours été retrouvés. Par ailleurs, j'ai montré que les locus TAS permet l'établissement de la répression des séquences qui s'y insèrent par la transmission de ses propriétés épigénétiques. Ce type de mécanisme pourrait être généralisé aux autres locus producteurs de piRNA. Ils assureraient ainsi la protection de l'intégrité du génome contre la transposition d'un nouvel élément qui arriverait dans une « trappe génomique » par l'effacement de son identité.

Mots clés : Telomere Associated Sequence, Eléments Transposables, Subtélomères, petits ARN non codants, piRNA, Drosophile